

PowerAnalysis.m Guide

Roger Strong
Harvard University

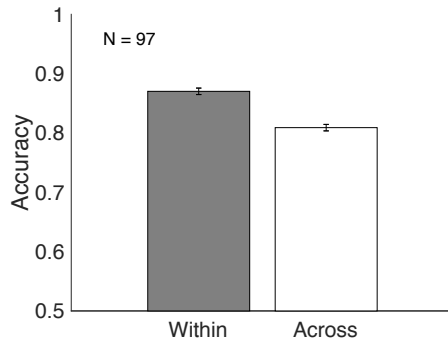
General Notes

- PowerAnalysis.m does most the work, and is called in the example scripts
- Key Components:
 - prefs.data:
 - either a #subjects (rows) x #conditions (columns) array, or a string file name of an excel or .csv file with data listed as #subjects x #conditions.
 - Data can be listed as either decimal (.5) or percentage (50), although you will get a warning for the later (as data will be converted to decimal)
 - If using excel or csv file, there should NOT be a header row
 - prefs.N_range
 - Range of number of participants to simulate. E.g., 10:10:50 will simulate with 10, 20, 30, 40, and 50 participants
 - prefs.trial_range
 - Range of number of trials per condition to simulate. E.g., 8:4:24 will simulate with 8, 12, 16, 20, and 24 trials per condition
 - prefs.alpha
 - p-value to use in power simulations
 - prefs.nSims
 - How many simulations to use for every participant/trial number combination. 10,000 is a decent estimate and runs pretty quickly, 100,000 is slower but a more stable estimate.
 - prefs.comps
 - Which comparisons to test for significance. Each row is a comparison, with the condition expected to be higher magnitude listed in the first column, and the condition expected to have lower magnitude in the second column. A study will be classified as “successful” only if all listed comparisons are significant (see examples).

Example 1

Pilot Data

- 97 subjects, 2 conditions
- Excel file is 97 rows x 2 columns



Exp1_Data.xlsx →

	A	B
1	0.8125	0.71875
2	0.84375	0.625
3	0.375	0.625
4	0.71875	0.6875
5	0.84375	0.875
6	0.5	0.5
7	0.96875	0.84375
8	0.75	0.71875
9	0.9375	0.78125
10	0.875	0.78125
11	0.84375	0.84375
12	0.625	0.71875
13	0.59375	0.625
14	0.90625	1
15	0.90625	0.9375
16	0.9375	0.78125
17	0.90625	0.6875
18	0.9375	0.875
19	0.84375	0.71875
20	0.4375	0.46875
21	1	0.90625
22	0.90625	0.90625
23	1	0.8125
24	0.875	0.75
25	0.5	0.46875
26	0.6875	0.65625
27	0.9375	0.9375
28	0.71875	0.8125
29	0.9375	0.9375
30	0.875	0.875
31	0.8375	0.78125

Power Analysis Settings

```
clear
%can either be your data as a sub * cond matrix,
% or name of an excel/csv file as str
prefs.data = 'Exp1_Data.xlsx';

%interval of N to simulate (e.g, 10-100 by 10)
prefs.N_range = 10:10:100;

%interval of trials per condition to simulate (e.g, 8-24 by 4)
prefs.trial_range = 8:4:24;

%p value to use in statistical test during simulation
prefs.alpha = .05;

%number of experiments to simulate per trial*N combination
%higher number of sims will give more stable/accurate power estimates,
%but will be slower. 10000 or 100000 is usually good
prefs.nSims = 10000;

%what comparisons do you want to make? Should be a comparison * 2 vector,
%with condition that should be larger on the left
%for example, if you expect condition 1 to be larger than condition 2, you
%should enter [1, 2];
prefs comps = [1, 2];

%Run Power Analysis with these settings
pow_results = PowerAnalysis(prefs);
```

File name as string (can also do data directly in matlab)

I decided to simulate N from 10-100 by 10

I decided to simulate trial number per condition from 10-100 by 10

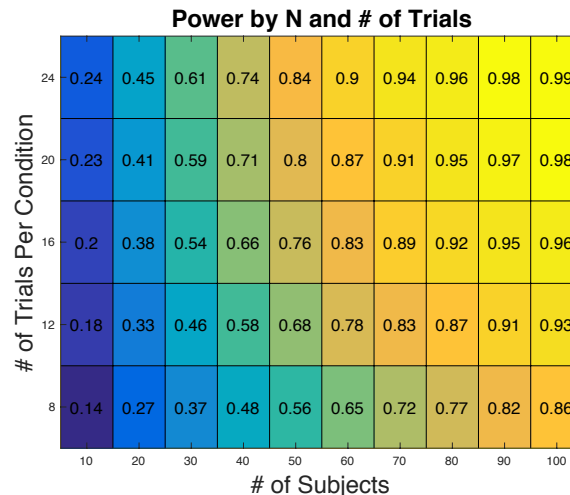
P-value of .05 used in simulation

10,000 sims per N x num_trials combo (sims per cell in output graph)

Only comparison I was interested in was condition 1 being larger than condition 2

Run power analysis using these settings

Power Analysis Output



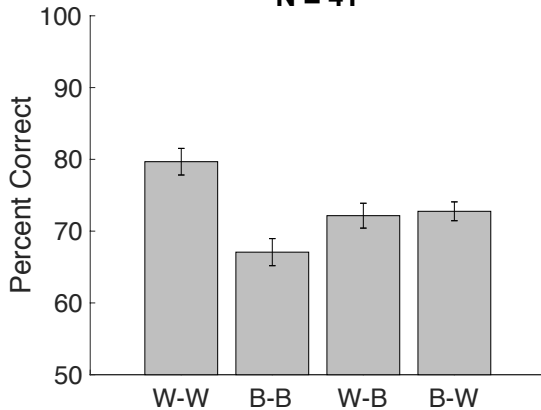
Simulated power for each N X number or trials per condition combo we specified in settings. Looking at this, I know I could achieve > 90% power by running 90 subjects with 12 trials per condition, for example

Example 2

Pilot Data

- 41 subjects, 4 conditions
- Excel file is 41 rows x 4 columns

N = 41



Exp2_Data.xlsx

	A	B	C	D
1	41.6666667	33.3333333	83.3333333	50
2	83.3333333	75	75	75
3	50	91.6666667	50	75
4	91.6666667	50	91.6666667	66.6666667
5	100	66.6666667	91.6666667	100
6	100	83.3333333	91.6666667	100
7	58.3333333	33.3333333	66.6666667	50
8	75	58.3333333	50	66.6666667
9	100	100	75	91.6666667
10	41.6666667	33.3333333	50	50
11	100	100	91.6666667	83.3333333
12	50	58.3333333	83.3333333	66.6666667
13	91.6666667	58.3333333	66.6666667	50
14	91.6666667	58.3333333	75	58.3333333
15	91.6666667	58.3333333	50	66.6666667
16	100	50	75	66.6666667
17	100	83.3333333	91.6666667	91.6666667
18	100	91.6666667	100	100

Data is in percent, so script will convert to decimal and give a warning that this has occurred.

Power Analysis Settings

```
clear
%can either be your data as a sub * cond matrix,
% or name of an excel/csv file as str
prefs.data = 'Exp2_Data.xlsx';

%interval of N to simulate (e.g, 50-300 by 25)
prefs.N_range = 50:25:300;

%interval of trials per condition to simulate (e.g, 8-20 by 4)
prefs.trial_range = 8:4:20;

%p value to use in statistical test during simulation
prefs.alpha = .05;

%number of experiments to simulate per trial*N combination
%higher number of sims will give more stable/accurate power estimates,
%but will be slower. 10000 or 100000 is usually good
prefs.nSims = 10000;

%what comparisons do you want to make? Should be a comparison * 2 vector,
%with condition that should be larger on the left
%for example, if you expect condition 1 to be larger than condition 2, you
%should enter [1, 2];
prefs comps = [1, 2
               1 3
               1 4
               3 2
               4 2];

%Run Power Analysis with these settings
pow_results = PowerAnalysis(prefs);
```

File name as string (can also do data directly in matlab).

I decided to simulate N from 50-300 by 25

I decided to simulate trial number per condition from 8-20 by 4

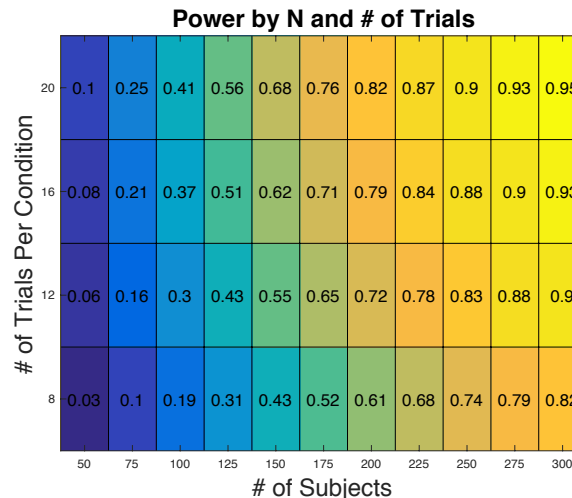
P-value of .05 used in simulation

10,000 sims per N x num_trials combo (sims per cell in output graph)

This time, I had 5 comparisons I am interested in. Specifically, I only want to call the study a "success" if condition 1 > 2, 1 > 3, 1 > 4, 3 > 2, and 4 > 2. Each comparison specified as a separate row.

Run power analysis using these settings

Power Analysis Output



Simulated power for each N X number of trials per condition combo we specified in settings. Looking at this, I know I could achieve > 90% power by running 300 subjects with 12 trials per condition, for example. Note that this is power for ALL 5 comparisons of interest being significant