Power Analysis

2021/5/29

Huanqing Wang

Outline

- Power
  - Sample size
  - Significance
  - Effect size
- Power Analysis
  - Univariate
  - MVPA

Power

Power is **the probability that we will correctly reject** the Null Hypothesis.

Alternatively, power is the probability that we will correctly get a small p-value (<.05)

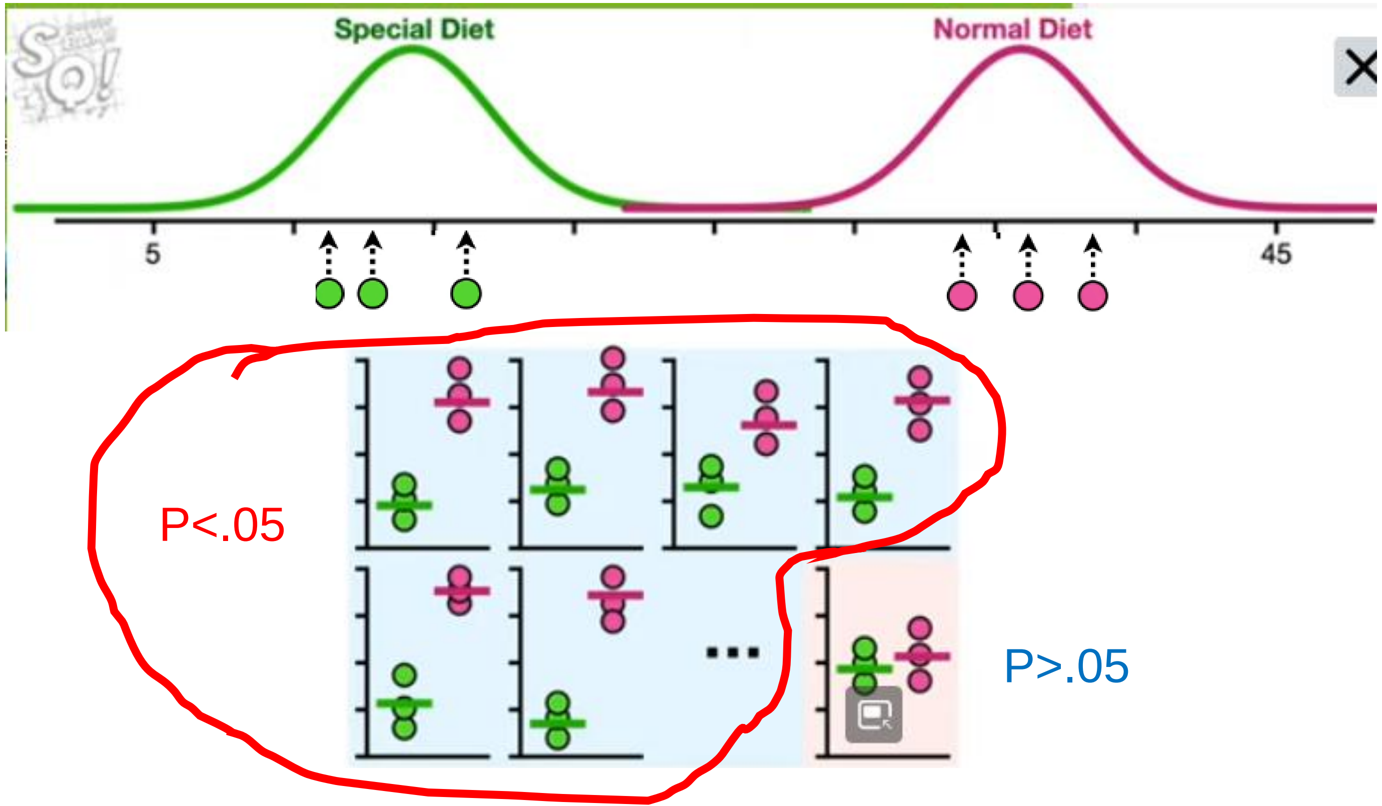
Power is **the probability that we will correctly reject** the Null Hypothesis.

Alternatively, power is the probability that we will correctly get a small p-value (<.05)

	Null is True	Null is False
Reject Null	α Type I error	$1-\beta$ (power) Correct Decision
Do not reject Null	$1-\alpha$ Correct Decision	β Type II error

Power is **the probability that we will correctly reject** the Null Hypothesis.

Alternatively, power is the probability that we will correctly get a small p-value (<.05)



Values related to Power

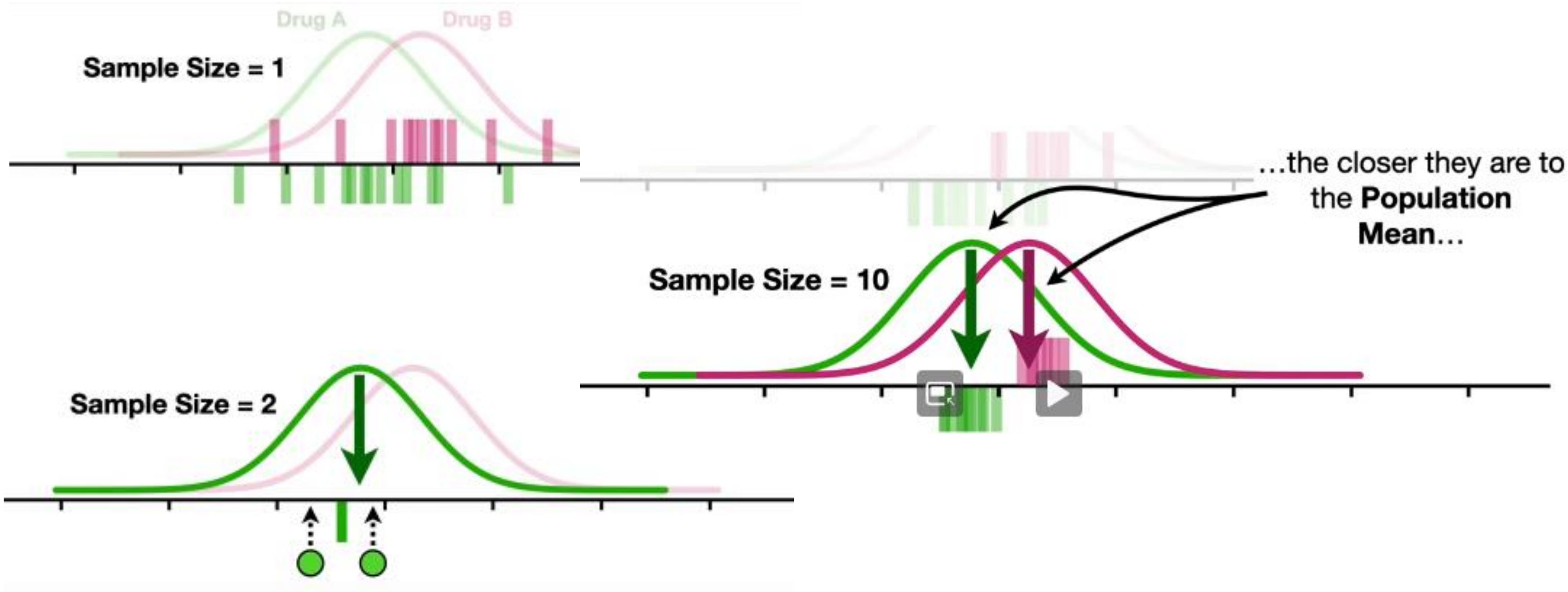
Sample size

Significance level (α)

Effect size

Values related to Power

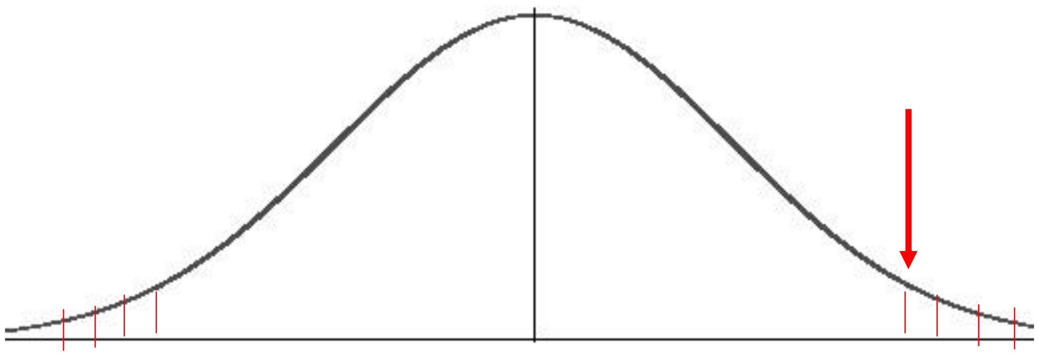
Sample size



Values related to Power

Significant level (p-value)

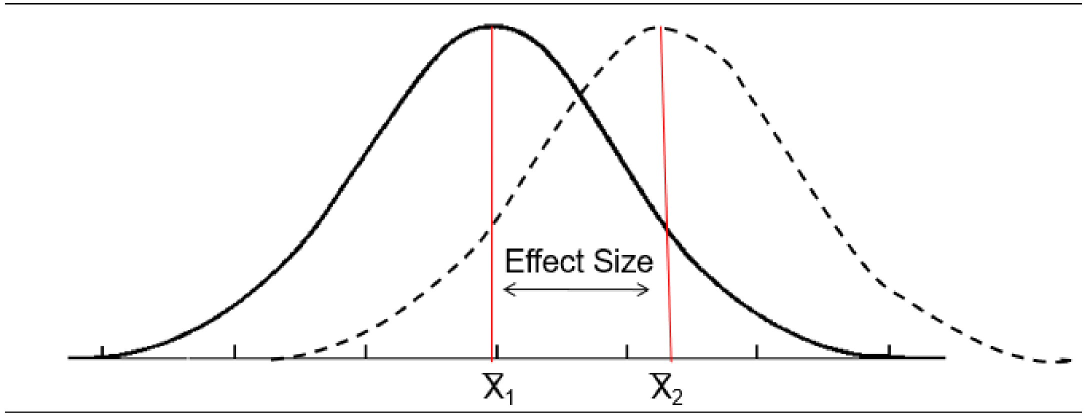
A p-value is the probability that random chance generated the data, or something else that is **equal or rarer.**



Values related to Power

Effect size

Effect is a measure of the **difference between two populations**

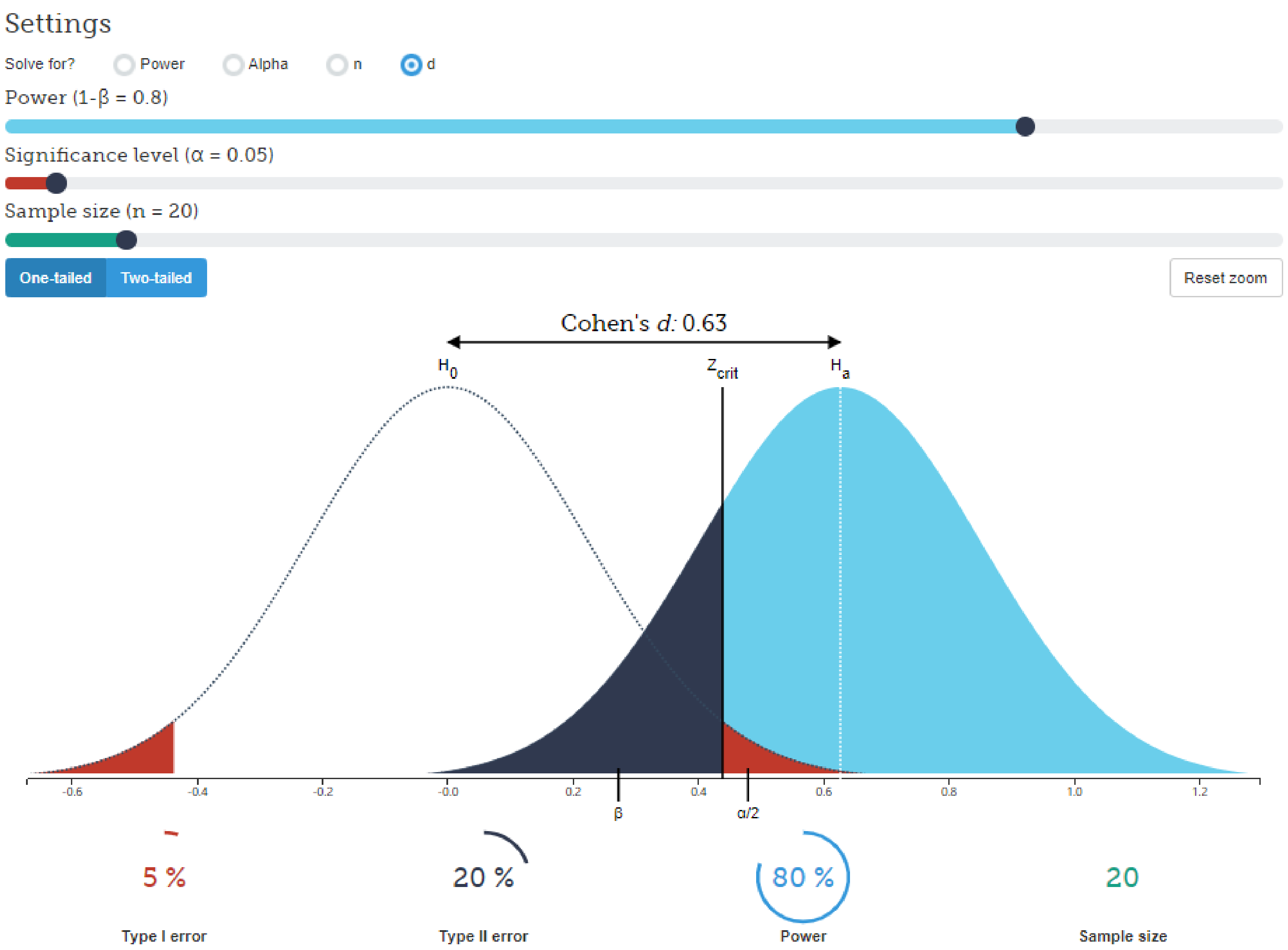


The bigger the effect size, the bigger the difference in the means.

Significance ≠ Large effect size

假设检验：检测处理效应的相对大小，效应量：处理效应的绝对大小

Relationships between them



Power Analysis

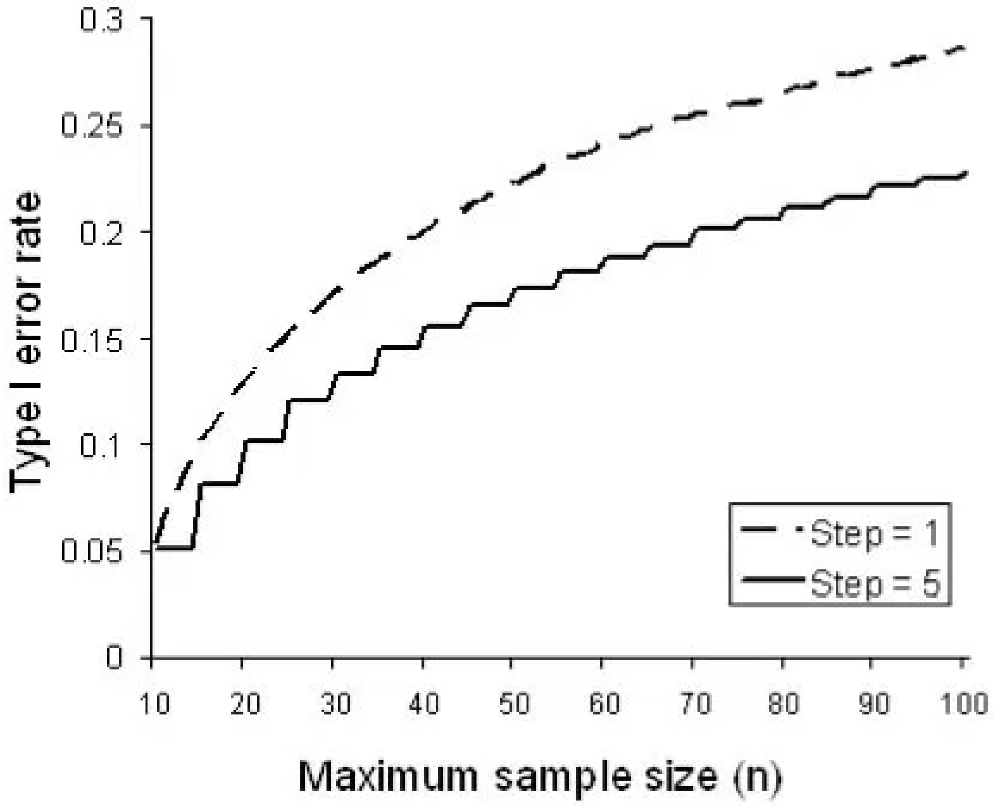
Data peeking

Imagine you collect 10 subjects and you perform an analysis. You see two possibilities:

1. You find the significant effects you were hypothesising:
Woohoo! This means I can stop my experiment, write my paper and publish my findings.

2. Not everything you expected is significant, but there is a trend:
You decide to add a few more subjects and then look at the data again.

Data peeking



Power Analysis

Power Analysis will tell us **how many measurements** we need to collect to have a good amount of Power.

If we use the **sample size** recommend by the **power analysis**, we will know that, regardless of the p-value, **we used enough data to make a good decision**.

Power Analysis

Steps

First: how much Power we want (a common value is **0.8**);

Second: To determine the threshold for significance (alpha α), common is **0.05**;

Lastly: Effect size;

Finally, use "statistics power calculator" to calculate needed Sample Size.

Power Analysis For Behavioral Experiment

G Power

确定实验设计和变量类型 → 选择统计方法（t-test、ANOVA、regression等）→ 选择分析类型（priori、post hoc等）→ 输入参数（α、1-β等）→ 输出结果（sample size、power等）

Power Analysis For Behavioral Experime

G Power

Independent t-test

Test family: t tests;

Statistical test: Means: Difference between two independent means (two groups);

Tail(s): Two;

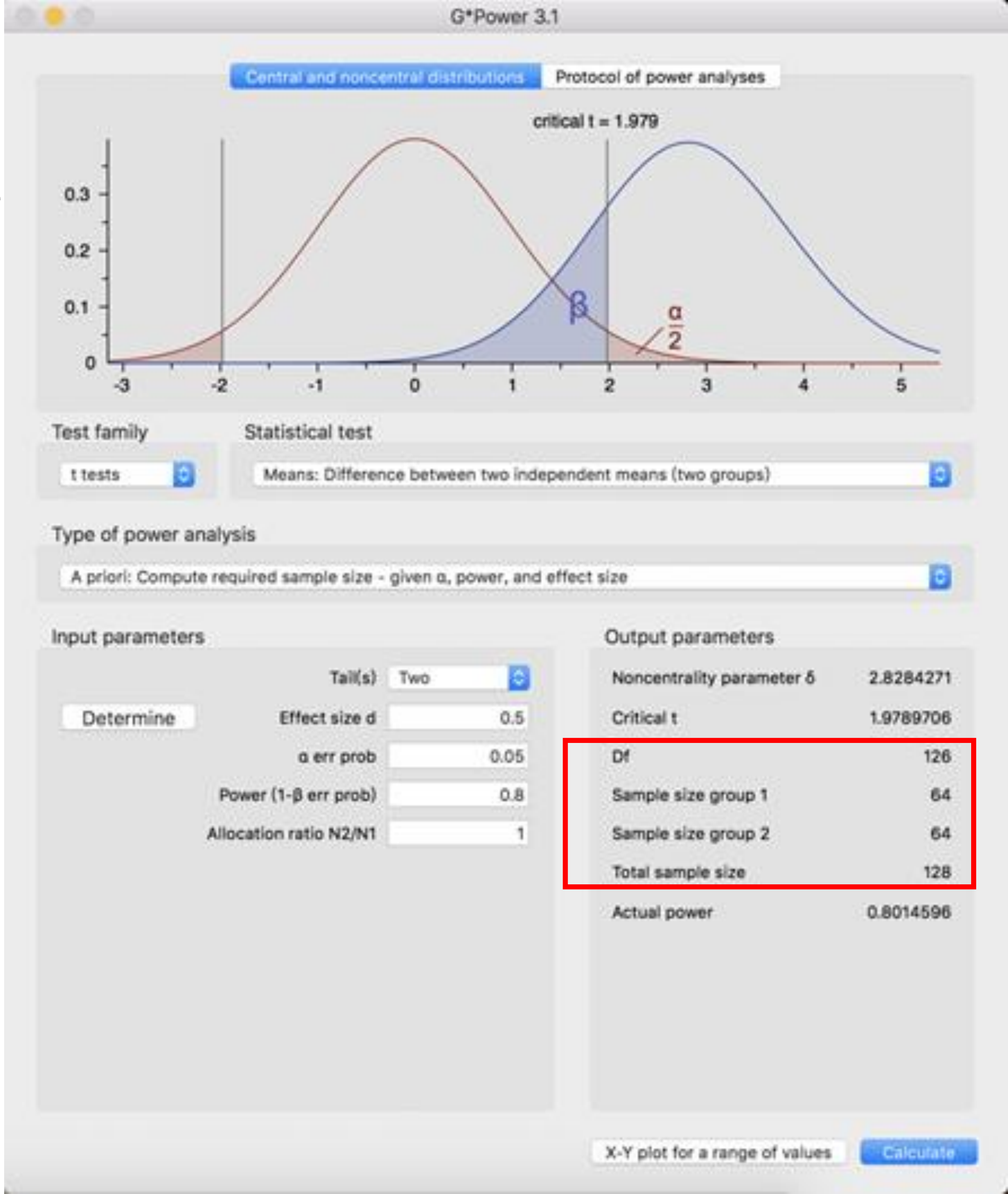
Effect size d: 0.5;

α err prob: 0.05;

Power: 0.8;

Allocation ratio N2/N1: 1

128人的样本量可以在显著性水平（α）小于等于0.05和统计检验力（1-β）为0.8时检测出处理间中等大小的主效应（d = 0.5）



Power Analysis For Behavioral Experime

两因素混合设计 2X3 mixed design

Between factors main effect

Test family: F tests;

Statistical test: ANOVA: Repeated measures, between factors;

Effect size f: 0.25;

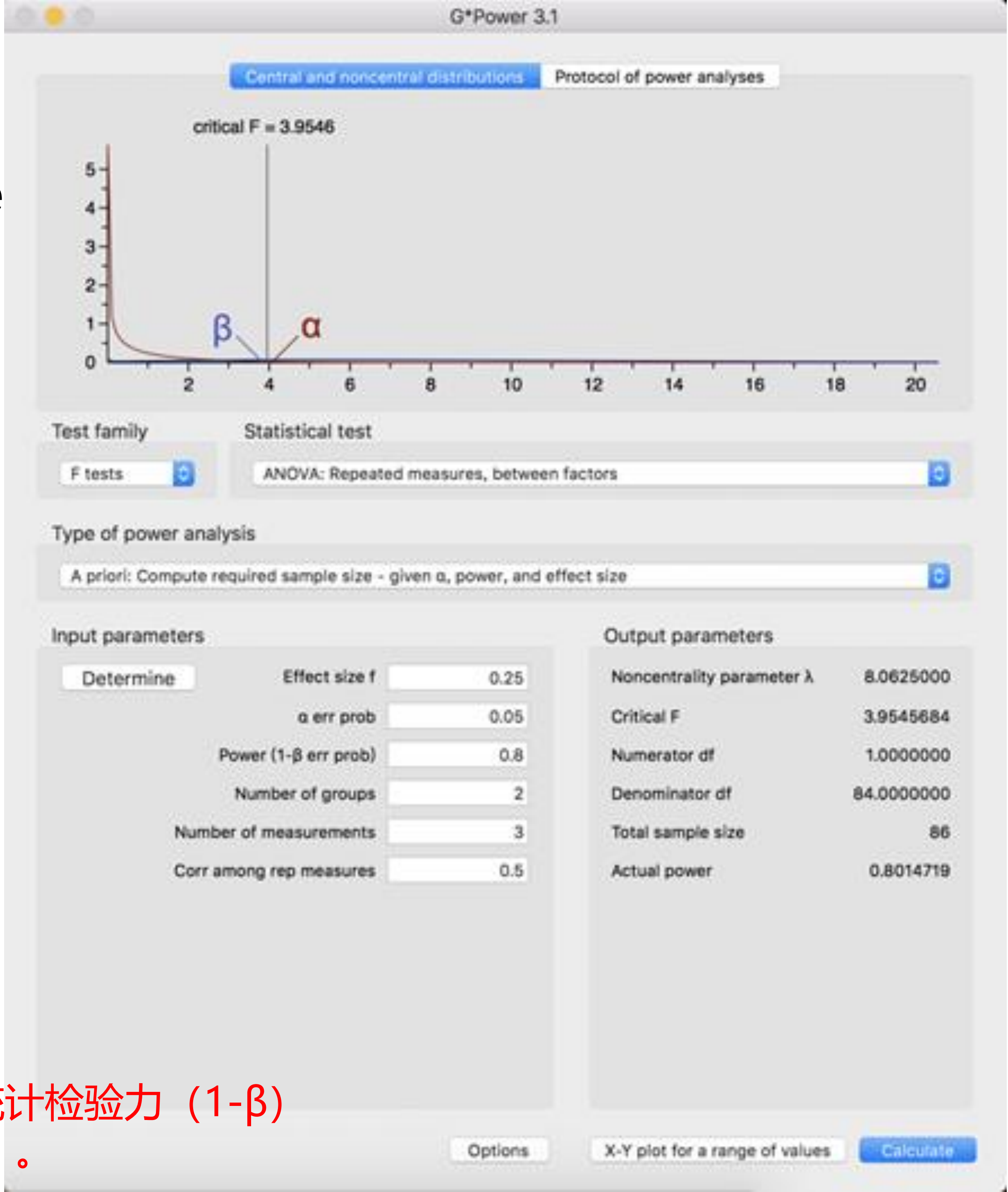
α err prob: 0.05;

Power: 0.8;

Number of groups: 2;

Number of measurements: 3;

Corr among rep measures: 0.5.



86人的样本量可以在显著性水平（α）小于等于0.05和统计检验力（1-β）为0.8时检测出被试间因素中等大小的主效应（f = 0.25）。

Power Analysis For Behavioral Experime

两因素混合设计 2X3 mixed design

Within factors main effect

Test family: F tests;

Statistical test: ANOVA: Repeated measures, within factors;

Effect size f: 0.25;

α err prob: 0.05;

Power: 0.8;

Number of groups: 2;

Number of measurements: 3;

Corr among rep measures: 0.5;

Nonsphericity correction ε = 1.

输出：28人的样本量可以在显著性水平（α）小于等于0.05和统计检验力（1-β）为0.8时检测出时间点中等大小的主效应（f = 0.25）。

Power Analysis For Behavioral Experime

两因素混合设计 2X3 mixed design

Interaction

Test family: F tests;

Statistical test: ANOVA: ANOVA: Repeated measures, within-between interaction;

Effect size f: 0.25;

α err prob: 0.05;

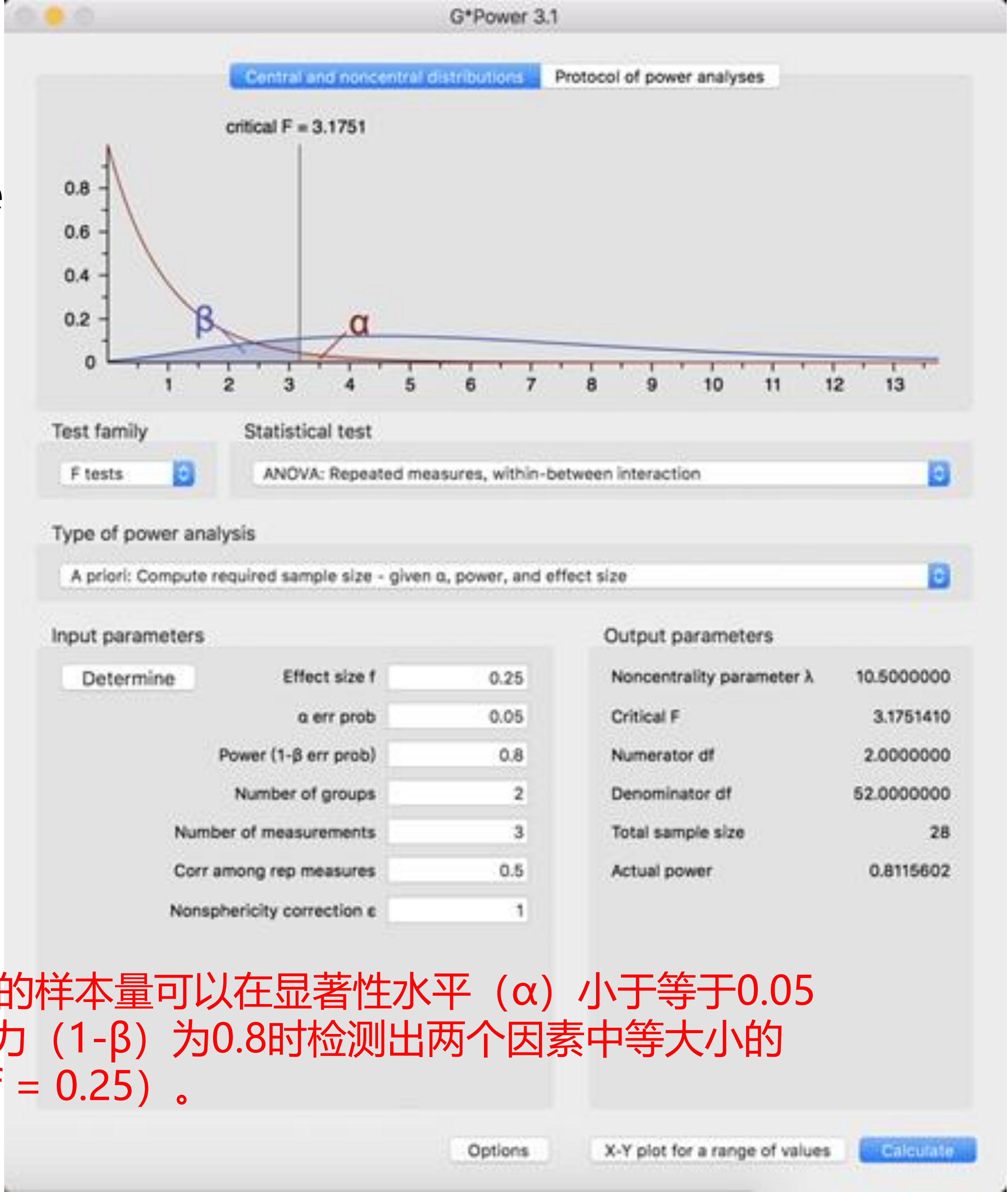
Power: 0.8;

Number of groups: 2;

Number of measurements: 3;

Corr among rep measures: 0.5;

Nonsphericity correction ε = 1.



输出：28人的样本量可以在显著性水平（α）小于等于0.05和统计检验力（1-β）为0.8时检测出两个因素中等大小的交互效应（f = 0.25）。

常用效应量系数解读表

效应量解读 (Interpretation)	d/d_z*	r*	r^2	η^2	f	f^2	Odds ratio**	AUC	z
	0.0	0.0	0.00	0.00	0.00	0.00	1.00	0.50	0.00
小效应量 (Small effect)	0.2	0.1	0.01	0.01	0.10	0.02	1.68	0.56	0.10
中效应量 (Medium effect)	0.5	0.3	0.09	0.06	0.25	0.15	3.47	0.64	0.35
大效应量 (Large effect)	0.8	0.5	0.25	0.14	0.40	0.35	6.71	0.71	0.40

Power Analysis For fMRI Experiment

Univariate

Neuropower

1 Pilot data

2 Data input

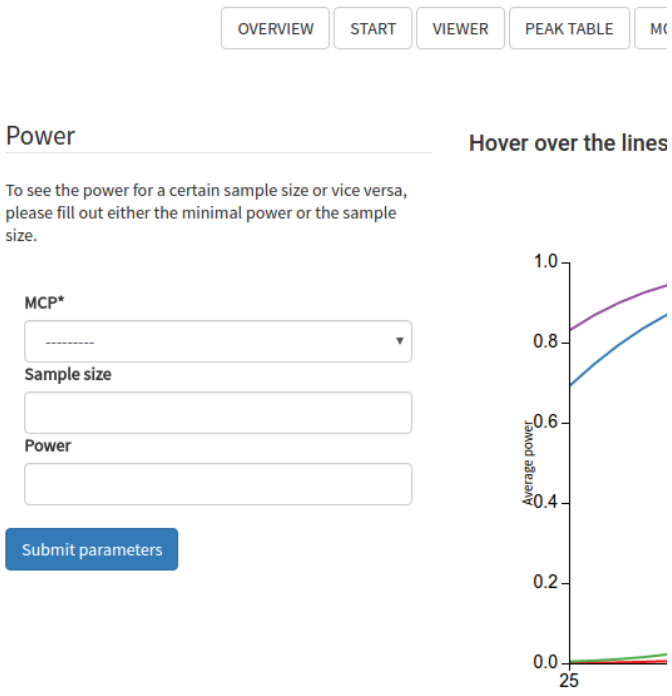
3 Peak table

4 Model Fit

5 Power Analysis

Power Analysis For fMRI Experiment

Univariate



OVERVIEW | START | VIEWER | PEAK TABLE | MODEL FIT | POWER CALCULATION | POWER TABLE | RESET

Sample Size	Random Field Theory	Bonferroni	Benjamini-Hochberg	Uncorrected
25	0.00	0.00	0.69	0.83
26	0.01	0.00	0.74	0.86
27	0.01	0.00	0.79	0.90
28	0.02	0.00	0.83	0.92
29	0.02	0.01	0.87	0.94
30	0.03	0.01	0.90	0.96
31	0.04	0.01	0.92	0.97
32	0.06	0.02	0.94	0.98
33	0.08	0.02	0.96	0.98
34	0.10	0.03	0.97	0.99
35	0.13	0.05	0.98	0.99
36	0.16	0.06	0.98	1.00
37	0.19	0.08	0.99	1.00
38	0.23	0.10	0.99	1.00

Power Analysis For fMRI Experiment

MVPA

run simulation

Effect size r = ? Correlation between behavioral similarity matrix and neural similarity matrix

Simulated activity in a single searchlight Create behavioral model

Created patterns of neural activity Add additional random noise on behavioral model

Added additional noise ~N(0, 10) to represent data from individual trials

Iteration On each iteration of the simulation, a particular participant number and trial (per stimuli) were set

Power Analysis

A Monte-Carlo simulation in MATLAB 7 was conducted to establish a design with adequate statistical power. We simulated a behavioral similarity matrix and a neural similarity matrix that were correlated at the estimated population effect size of $r = 0.15$. This effect size was thought to be reasonable based on previous work (e.g., ref. 13). To generate the behavioral similarity matrix, we simulated activity in a single searchlight. That searchlight consisted of 200 voxels and 60 separate patterns of activity, to represent each of the mental states. The simulated "activity" within the voxels should be normally distributed ($M = 0$, SD = 1) as an approximation for the t-values used in the actual analysis. The 60×60 correlation matrix produced by this searchlight was considered the behavioral model. We created patterns of neural activity within the simulated searchlight by taking the 200 voxel-by-60 state matrix used to generate the behavioral model and adding additional random noise: $\sim N(0, 2.4)$. When these neural patterns were correlated with one another, the resulting neural similarity matrix consistently correlated with the corresponding behavioral matrix at approximately $r = 0.15$. Because this neural pattern matrix reflects experiment-level data, we added additional noise $\sim N(0, 10)$ to represent data from individual trials.

On each iteration of the simulation, a particular participant number and trial (per mental state) were set. Trial-wise neural patterns of searchlight activity were generated for each participant and averaged to produce a single pattern for each participant. These patterns were then converted to similarity matrices and correlated with the overall behavioral similarity matrix. The resulting r values were R-to-z transformed and entered into a t test across simulated participants. The result of this t test was tabulated to estimate power. Participant numbers between 2 and 30 and item numbers between 2 and 20 were simulated, with 100 simulation iterations at each combination of these parameters. These simulations indicated that 20 participants with 16 trials per mental state should be adequate to ensure 95% voxelwise statistical power at an uncorrected threshold of $P < 0.001$.