Power!

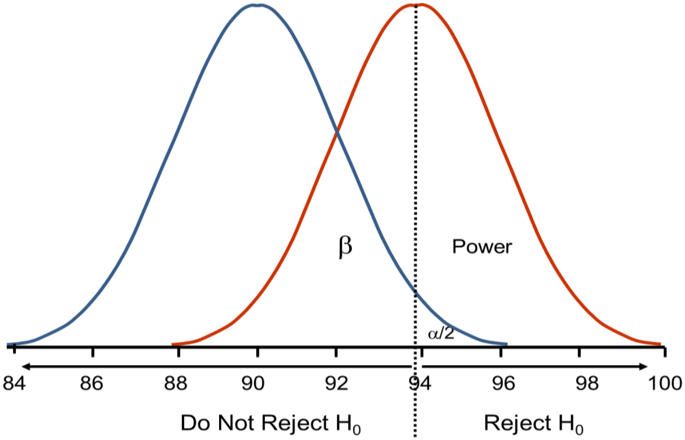
Kevin M. Esterling

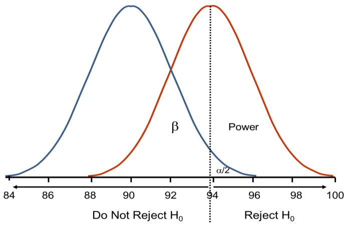
UC Riverside

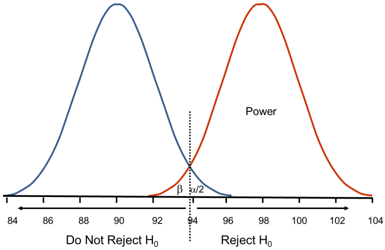
November 2, 2017

Hypothesis testing and power

- In null hypothesis statistical testing (NHST) we test $H_0 : \tau = 0$ against $H_1 : \tau \neq 0$
- Two types of research error in NHST:
- **Type I error:** falsely reject null when it is true
  - Controlled by the significance level or "size" of the test
  - At $\alpha = 0.05$ only 5 times in 100 would you observe $p < 0.05$
- **Type II error:** falsely reject the alternative when it is true
  - Power is the percent of time alternative is accepted when true
  - Usual symbol is κ (sometimes $\beta = 1 - \kappa$)
  - Higher power is better – more likely to detect an effect for a given significance level


β
Power
α/2
84
86
88
90
92
94
96
98
100
Do Not Reject H0
Reject H0


β
Power
α/2
84
86
88
90
92
94
96
98
100
Do Not Reject H_0
Reject H_0


Power
β
α/2
84
86
88
90
92
94
96
98
100
102
104
Do Not Reject H_0
Reject H_0

Sample size/ Power

Power is the probability that you can reject a false null hypothesis.
Several elements of your study determine the power.

- Elements typically not under your control:
  - Effect size (β)
  - Variability in the outcomes (σ^2)
- Elements typically under your control:
  - Sample size (N)
  - Power requirement (κ)
  - Significance level (α)
  - Proportion assigned to treatment v. control (P)
  - Including covariates that are correlated with outcomes (X)

t units

To test a hypothesis, and to describe a study's power, we need to rescale test statistic into "t" units

$$t = \frac{ATE}{SE(ATE)} \tag{1}$$

Estimate the ATE using regression

In order to run the regression to estimate the ATE (i.e., β), we assume,

$$Y_i = \alpha + \beta Z_i + \epsilon_i \tag{2a}$$

$$\epsilon_i \sim N(0; \sigma^2) \tag{2b}$$

Then the standard error of the ATE is

$$\sqrt{\frac{1}{P(1-P)}\frac{\sigma^2}{N}} \tag{3}$$

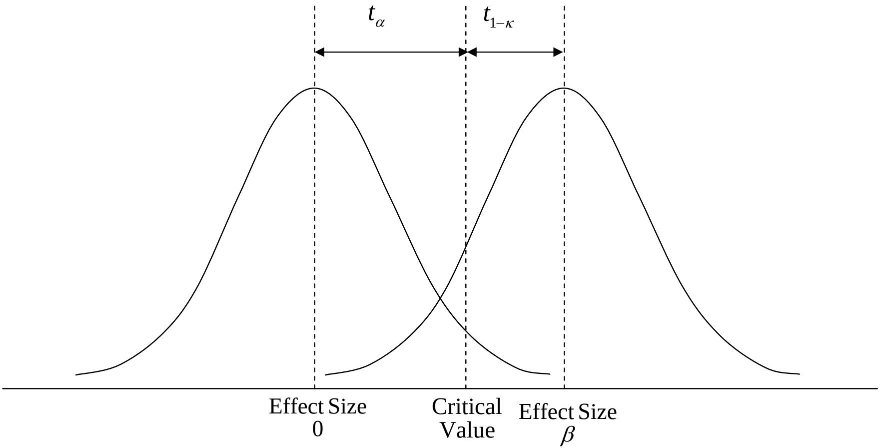
Convert t units to β units

We can convert back and forth between t and β units

$$t = \frac{\beta}{SE(\beta)} \tag{4a}$$

$$\beta = t \times SE(\beta) \tag{4b}$$

Let's derive the basic power formula ...


t_α
$t_{1-\kappa}$
Effect Size
0
Critical
Value
Effect Size
β

Deriving the basic power formula

We need to find the area to the right of the critical value for the alternative hypothesis sampling distribution

$$\beta = t \times SE(\beta) \tag{5a}$$
$$\beta = (t_\alpha + t_{(1-\kappa)}) \times SE(\beta) \tag{5b}$$
$$\frac{\beta}{SE(\beta)} - t_\alpha = t_{(1-\kappa)} \tag{5c}$$

. . . Then look up the area to the right of $t_{(1-\kappa)}$

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.

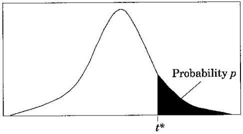


Table B *t* distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Minimum detectable effect

- Assume estimate ATE with OLS, we can state a MDE by setting $\alpha = 0.05$, $\kappa = 0.80$, sample size of N, proportion assigned to treatment group P, and estimate $\widehat{\sigma^2}$
- The minimum detectable effect (MDE) is

$$MDE = [t_\alpha + t_{(1-\kappa)}] \times \sqrt{\frac{1}{P(1-P)} \frac{\sigma^2}{N}} \tag{6a}$$

$$= \text{“t-distance”} \times SE(\beta) \tag{6b}$$

Including covariates in the analysis can improve your power

Covariates can reduce the MSE (σ^2) of your regression

$$Y_i = \alpha + Z\beta + (\epsilon_i); \epsilon_i \sim N(0; \sigma^2) \quad (7a)$$

$$= \alpha + Z\beta_1 + (X\beta_2 + \epsilon_i^*); \epsilon_i^* \sim N(0; (\sigma^*)^2) \quad (7b)$$

If $\beta_2 \neq 0$, then $(\sigma^*)^2 < \sigma^2$

Level of randomization/assignment

- Individual
- Group or cluster
  - Classroom
  - School
  - School district
  - Village/city

Why clustered designs?

- Some treatments can only be at cluster level (Classroom, Lab, Small group)
- Need to account for interference/SUTVA violation

Power in clustered designs

- Within-group correlation in outcomes reduces power
  - Common background, information, omitted factors
  - More power the more unrelated the people in the group
- ρ is the *intra-cluster correlation* (ICC)
- Need to adjust power calculation with a design effect. For a given sample size, clustering in groups of size m, the MDE increases by $\sqrt{1+\rho(n-1)}$
- Usually the number of groups matters more than the number of individuals
- Issues:
  - Individual randomization gives bigger sample, more power
  - Beware of nesting even when assigning at individual level

Minimum detectable effect for a clustered design

Assume J groups each of size n

$$\begin{aligned} ClusterMDE =& (t_{\alpha} + t_{(1-\kappa)}) \times \sqrt{\rho + \frac{1-\rho}{n}} \frac{\sigma}{\sqrt{P(1-P)J}} && \text{(8a)} \\ =& \text{“t-distance”} \times SE(\beta) && \text{(8b)} \end{aligned}$$

Note the Cluster MDE decreases more rapidly in J than in n

Resources for Power

- The Randomization Toolkit
- Alex Coppock's "10 Things to Know about Power" on EGAP

Calculating power requirements

- Decide the null hypothesis
- Set a significance level
- Set power requirement
- Decide sample size and proportion allocated to treatment, estimate outcome variance, include covariates
- Calculate MDE
- See if study will detect a minimum effect size you need

Ethical considerations

- Consider ethics and fairness of using other people's time for your own research
- Consider the costs and funding sources; opportunity cost for wasted resources
- Be clear about the time requirements for doing a careful study and how much of everyone's time you will be using
- Be transparent in conducting and reporting your analysis