Sample Size and Power

Henry Glick
www.uphs.upenn.edu/dgimhsr

May 12, 2017

Goals of Sample Size and Power Analysis

- Sample size calculation
  - Given a desired alpha (α) and beta (β), proactively manages probability of saying a difference exists when none does
    - Type 1 error; False positive; alpha; confidence
- Power Analysis
  - Given a desired alpha and a known sample size, proactively understand probability of saying no difference exists when one does
    - Type 2 error; False negative; (1-Beta); power
- "Experiment has an 80% chance (power) of concluding with 95% confidence (alpha) that therapies differ"

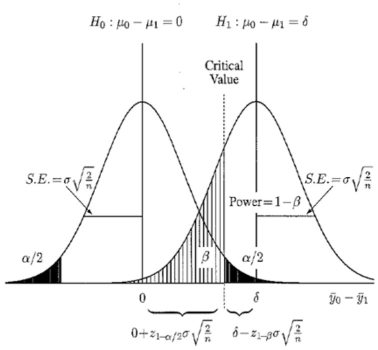
Other Sample Size Traditions

- Sample size approach described here comes out of frequentist statistical tradition
- Other approaches (cost-effectiveness literature) include:
  - Bayesian (O'Hagan and Stevens)
  - Value of information (Koerkamp et al.)
  - Opportunity cost (Gafni et al.)
  - Decision model (Willan and O'Brien)

Sample Size Calculation

- Specify parameters (e.g., expected difference (Δ) and standard deviation (SD)), alpha, and beta and identify sample size per group (N)
  - Assumptions of equivalent N/group or equivalent SD can be relaxed (see later equations)

Common Sample Size/Power Figure



www.vanbelle.org/chapters%5Cwebchapter2.pdf

Common Equations

Consider an α-level test of H_0: $\mu_1 = \mu_2$ against the one-sided alternative H_A: $\mu_1 > \mu_2$. The critical region for this test is

$$\bar{X} - \bar{Y} > z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \qquad (2.2.20)$$

If a power of $1 - \beta$ is required at $\mu_1 - \mu_2 = \delta(>0)$, n_1 and n_2 must satisfy

$$P\left(\bar{X} - \bar{Y} > z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \,\middle|\, \mu_1 - \mu_2 = \delta\right) = 1 - \beta. \qquad (2.2.21)$$

Thus,

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \left\{\frac{\delta}{z_\alpha + z_\beta}\right\}^2. \qquad (2.2.22)$$

To get an unique solution to (2.2.22), one takes $n_1 = kn_2$ and gets

$$n_2^* = \frac{\frac{\sigma_1^2}{k} + \sigma_2^2}{\left(\frac{\delta}{z_\alpha + z_\beta}\right)^2}, \qquad n_1^* = kn_2^*. \qquad (2.2.23)$$

Desu and Raghavarao, Sample Size Methoology

Not What I Am Going to Talk About

- If the graph and equations are enlightening, you may know everything you need to know
- I don't find them particularly helpful
- Hoping to provide some intuitions about power and sample size

Start with Naïve Sample Size Calculation

- Suppose we guestimate that our experiment will yield
  - Difference of 10
  - Standard deviation/group of 18.038
- If we run experiment and our data exactly match these assumptions, we need a standard error (SE) of 5.102 for a confidence interval where one confidence limit = 0 (i.e., p = 0.05)

  (10 – 1.96*5.102 = 0) (where 1.96 = z_α)
- Given we know SD, needed SE, and formula for SE – which includes sample size/group and SD – can make an initial naïve sample size estimate by substituting equation for SE and solving for N

Naïve Estimate of 25/Group

- CL = $\Delta - z_\alpha$ SE = 0
- Δ / z_α = SE
- SE= $(2 * (SD / N^{0.5})^2)^{0.5}$
- $\Delta / z_\alpha = (2 * (SD / N^{0.5})^2)^{0.5}$
- Solving for N/group yields:

$$N_{naive} = \frac{2\, z_\alpha^{\ 2}\ sd^2}{\Delta^2}$$

- Assuming Δ=10; sd=18.038; z=1.96: N=25/group

Effective α

- Using z_α = 1.96
- Assuming = 0.05; two groups; and no adjustment for multiple tests
- Results don't depend on assumptions
- What matters is that you use correct z_α after adjustment when calculating both CI and p-values and when calculating N/group
  - e.g., if after adjustment you decide that α = 0.05 when z_α = 2.2414, naïve sample size = 33/group

$$2 * 2.2414^2 * 18.038^2 / 10^2 = 32.69$$

Simulation to Test 25/Group

- Repeatedly draw 100,000 sets of 2 samples from normal distributions with:
  - 25/group
  - Means of 100 in group 1 and 110 in group 2 (Δ=10)
  - Common SD of 18.038/group
- Basic Stata command:

```
drawnorm c1 c2, m(100 110) sd(18.038) n(25)
```

WHAT DO WE EXPECT WILL HAPPEN?

Expectations

- Group 1 mean ≈ 100
- Group 2 mean ≈ 110
- Mean of each SD ≈ 13.038
- Δ ≈ 10
- SE ≈ 5.102
- % significant ???

Simulation Results

Number of draws	100,000
mean, Sample 1	99.9999
SD, Sample 1	18.0327
mean, Sample 2	109.9877
SD, Sample 2	18.0399
ΔC	9.9878
SE, ΔC	5.1280
p < 0.05, %	49.5

example1822.dta

Sources of (Lack of) Significance

Number of draws	100,000
% > 10	49.7
% < 5.102	48.7

- Significance
  - Δ>10; SE<5.102
  - Δ>10; SE>5.102; Δ/SE>1.96
  - Δ<10; SE<5.102; Δ/SE>1.96
- Lack of significance
  - Δ<10; SE>5.102
  - Δ<10; SE<5.102; Δ/SE<1.96
  - Δ>10; SE>5.102; Δ/SE<1.96

My Expectations

ΔC<10; SE>5.1		ΔC<10; SE<5.1		ΔC>10; SE>5.1		ΔC>10; SE<5.1	
p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05
25,000	0	12,500	12,500	12,500	12,500	0	25,000
0%		50%		50%		100%	

- 50% significance due to
  - ~25% of time Δ > 10 and SE < 5.102
  - ~12.5% of time Δ > 10, SE > 5.102, Δ/SE > 1.96
  - ~12.5% of time Δ < 10, SE < 5.102, Δ/SE > 1.96

Observed

ΔC<10; SE>5.1		ΔC<10; SE<5.1		ΔC>10; SE>5.1		ΔC>10; SE<5.1	
p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05
25,694	0	21,599	3005	3207	22,379	0	24,116
0%		12.2%		87.5%		100%	

- 49.5% significance due to
  - ~25% of time Δ > 10 and SE < 5.102
  - ~22% of time Δ > 10, SE > 5.102, but Δ/SE > 1.96
  - ~3% of time Δ < 10, SE < 5.102, but Δ/SE > 1.96

Are We Satisfied with Naïve Sample Size Estimate?

- Only if we are willing to live with designing experiments in which we are likely to detect a significant difference 50% of time

WHAT WENT WRONG?

What Went Wrong?

- Did not account for fact that we cannot expect Δ and SE to equal 10 and 5.102 in repeated experiments
  - In simulation approximately 50% of time Δ and SE were above 10 and 5.102 and 50% of time Δ and SE were below
- Need to increase sample size so that expected SE is smaller than 5.102, such that we increase likelihood of:
  - $\Delta > 10$ and $SE < 5.102$ (primary mechanism)
  - $\Delta < 10$, $SE < 5.102$, $\Delta/SE > 1.96$ (secondary mechanism)

Suppose We Simulated 51/Group, Not 25

- Repeatedly draw 100,000 sets of 2 samples from normal distributions with:
  - 51/group
  - Means of 100 in group 1 and 110 in group 2 (Δ=10)
  - Common SD of 18.038/group

WHAT DO WE EXPECT WILL HAPPEN?

Expectations

- Group 1 mean ≈ 100
- Group 2 mean ≈ 110
- Mean of each SD ≈ 13.038
- $\Delta \approx 10$
- SE < 5.102
- % significant > 50%

Simulation Results

Number of draws	100,000
mean, Sample 1	99.9777
SD, Sample 1	18.0349
mean, Sample 2	110.007
SD, Sample 2	18.0294
ΔC	10.0294
SE, ΔC	3.5804
$p < 0.05$, %	79.7
% > 10	50.45
% < 5.102	100

example218126.dta

79.4% Significant

ΔC<10; SE>5.1		ΔC<10; SE<5.1		ΔC>10; SE>5.1		ΔC>10; SE<5.1	
p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05
0	0	20,298	29,256	0	0	0	50,446
NA		59.0%		NA		100%	

- 79.4% significance due to
  - ~50.4% of time Δ > 10 and SE < 5.102
  - ~29% of time Δ < 10, SE < 5.102, but Δ/SE > 1.96
- Further increases in sample size gain power solely from increasing proportion of time Δ < 10, SE < 5.102, but Δ/SE > 1.96

Shifting Sources of Significance

n = 5/G; power = 0.14

ΔC<10; SE>5.1		ΔC<10; SE<5.1		ΔC>10; SE>5.1		ΔC>10; SE<5.1	
p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05
49,553	0	298	25	33,498	16,308	0	318
0%		0.77		33%		100%	

n = 25/G; power = 0.5

ΔC<10; SE>5.1		ΔC<10; SE<5.1		ΔC>10; SE>5.1		ΔC>10; SE<5.1	
p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05
25,649	0	21,599	3005	3207	22,379	0	24,116
0%		12.2%		87.5%		100%	

n = 51/G; power = 0.7994

ΔC<10; SE>5.1		ΔC<10; SE<5.1		ΔC>10; SE>5.1		ΔC>10; SE<5.1	
p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05	p>0.05	p<0.05
0	0	20,298	29,256	0	0	0	50,466
NA		59.0%		NA		100%	

How Did We Come Up With 51/Group?

- Expand naïve sample size equation to include z_β (1.96 + 0.84 = 2.80)

$$N = \frac{2\,(z_\alpha + z_\beta)^2 \;\; sd^2}{\Delta^2}$$

- Target SE = 10/2.8 = 3.571 (3.58 in simulation)
- $N = 2 * 2.8^2 * 18.038^2 / 10^2 = 51$/group
- Power is typically treated as 1-tailed
  - If $z_\beta = 0$, power = 50% (equivalent to naïve sample size); if $z_\beta = 0.84$, power = 80%; if $z_\beta = 1.28$, power = 90%; if $z_\beta = 1.64$, power = 0.95

Stata sampsi Command for Sample Size

sampsi 100 110, sd(18.038) p(.8)

alpha	=	0.0500 (two-sided)
power	=	0.8000
m1	=	100
m2	=	110
sd1	=	18.038
sd2	=	18.038
n2/n1	=	1.00

Estimated required sample sizes:

n1	=	52
n2	=	52

Sample Size and Power

- Sample size and power mirror images of one another
- As previously noted, when estimating a sample size, specify parameters, alpha, and beta and identify number needed per group
- Power Analysis: specify parameters, alpha and sample size and identify power to detect a difference
- If we calculate sample size with beta=0.8 and determine 100 patients are needed per group, except for rounding, when we calculate power given 100 patients per group will see it equals 0.8

Power Equation, Continuous Variables

$$z_\beta = \sqrt{\frac{N\,\Delta^2}{2\ sd^2}} - z_\alpha$$

- Equation yields z_β
- Power identified from z table
  - In stata: normal(z_β)
- Assumes common N and common SD for each group

Power Equation, Continuous Variables

$$z_\beta = \sqrt{\frac{N\,\Delta^2}{2\ sd^2}} - z_\alpha = z_{cl=0} - z_\alpha$$

where $z_{cl=0}$ represents z-score that yields a CI for which one of CL = 0

- If know one of 90% CL equals 0 and want 95% confidence, know we have 38% power
  normal(1.645-1.96) = -0.376
- Works approximately for other types of contrasts as well
  - e.g., power for odds ratios
    - If one of the 90% CL equals 1 and want 95% confidence, know we have ≈ 38% power

Relaxing Some Assumptions

- Relaxing equal Ns in both groups:

$$N_2 = r\,N_1$$

$$z_\beta = \sqrt{\frac{r\,N_1\,\Delta^2}{(1+r)\ sd^2}} - z_\alpha$$

- Relaxing equal SDs in both groups

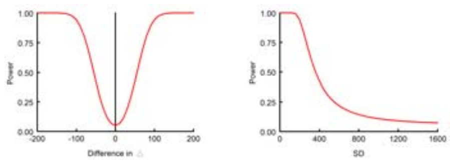
$$z_\beta = \sqrt{\frac{N\,\Delta^2}{sd_1^2 + sd_2^2}} - z_\alpha$$

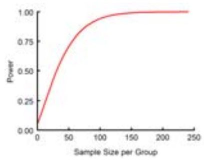
Stata sampsi Command for Power

```
sampsi 100 110, sd(18.038) n(52)

        alpha  =  0.0500 (two-sided)
           m1  =  100
           m2  =  110
          sd1  =  18.038
          sd2  =  18.038
           n1  =  52
           n2  =  52
        n2/n1  =  1.00
   Estimated power:
        power  =  0.8070
```

Patterns of Power For Δ, SD, and N *

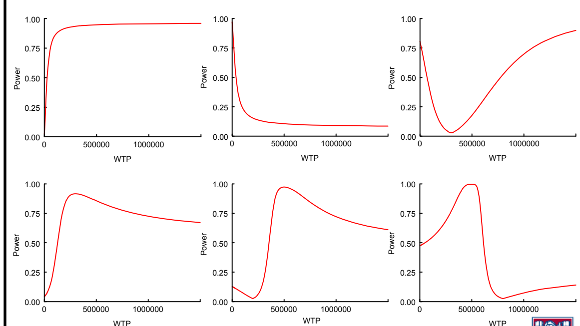




- α defines minimum power
  - e.g., if α = 0.05, can never have less than 5% power
  - Equivalent to "expect 1 in 20 FP if using α = 0.05"

* Unless otherwise noted, M1=200; M2=300; SD=200; N=100/ group

Six Power Patterns for Cost-Effectiveness Ratios



Glick HA. Sample Size and Power for Cost-Effectiveness Analyis, Part 1)

Incomplete Data (Drop Out)

- Derived sample size estimates are appropriate if we always have complete data
- If anticipate 10% with incomplete data, will want to divide sample size estimates by 0.9 to obtain "nominal" sample size from which "effective" sample size is derived