BIOS 662 Fall 2018

Power and Sample Size, Part III

David Couper, Ph.D.

david_couper@unc.edu

or

couper@bios.unc.edu

https://sakai.unc.edu/portal

Power and Sample Size

- Many of the sample size/power formulas assume a balanced design (exception: case-control)
- How do we generalize to unbalanced designs?
- For example, consider a two-sample t test with a continuous outcome

Two Sample t Test

- Assume normality and homogeneity of variance

$$\bar{Y}_i \sim N(\mu_i, \sigma^2/N_i) \quad \text{for} \quad i = 1, 2$$

- Under $H_0 : \mu_1 - \mu_2 = 0$

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{s_p\sqrt{1/N_1 + 1/N_2}} \sim t_{N_1+N_2-2}$$

- Under $H_A : \mu_1 - \mu_2 = \delta_A > 0$

$$\bar{Y}_1 - \bar{Y}_2 \sim N(\delta_A, \sigma^2(1/N_1 + 1/N_2))$$

implying

$$\frac{\bar{Y}_1 - \bar{Y}_2}{\sigma\sqrt{1/N_1 + 1/N_2}} \sim N\left(\frac{\delta_A}{\sigma\sqrt{1/N_1 + 1/N_2}}, 1\right)$$

Two Sample t Test

- Note

$$\frac{(N_1-1)s_1^2+(N_2-1)s_2^2}{\sigma^2} \sim \chi^2_{N_1+N_2-2}$$

- Therefore

$$\frac{(\bar{Y}_1-\bar{Y}_2)/(\sigma\sqrt{1/N_1+1/N_2})}{\sqrt{\frac{(N_1-1)s_1^2+(N_2-1)s_2^2}{\sigma^2(N_1+N_2-2)}}} \sim t_{N_1+N_2-2,\,\delta_A/(\sigma\sqrt{1/N_1+1/N_2})}$$

- Equivalently

$$\frac{\bar{Y}_1-\bar{Y}_2}{s_p\sqrt{1/N_1+1/N_2}} \sim t_{N_1+N_2-2,\,\delta_A/(\sigma\sqrt{1/N_1+1/N_2})}$$

Two Sample t Test

- Given N_1, N_2, α and δ_A, power is $\Pr[T > t_{1-\alpha/2}]$ where

$$T \sim t_{N_1+N_2-2,\, \delta_A/(\sigma\sqrt{1/N_1+1/N_2})}$$

- For example, suppose $N_1 = 10$, $N_2 = 20$, $\alpha = 0.05$, $\delta_A = 15$ and $\sigma = 25$

- R

```
> 1-pt(qt(0.975,28),28,15/(25*sqrt(1/10+1/20)))
[1] 0.3214083
```

Two Sample t Test

- SAS

```
proc power;
   twosamplemeans
   meandiff = 15
   groupns  = 10|20
   stddev   = 25
   power    = .;

Two-sample t Test for Mean Difference

      Fixed Scenario Elements

Distribution                  Normal
Method                         Exact
Mean Difference                   15
Standard Deviation                25
Group 1 Sample Size               10
Group 2 Sample Size               20

Power

0.322
```

Drop-outs and Loss to Follow-up

- Drop-out:
  - one who terminates involvement in an activity
  - a subject who withdraws from a trial or follow-up study by an announced unwillingness to continue to submit to required procedures
  - a subject who refuses or stops taking the assigned treatment
  - a subject who misses a scheduled visit
- A drop-out may drop out of treatment and/or out of follow-up

Drop-outs and Loss to Follow-up

- Drop-in:
  - a subject enrolled in a clinical trial who receives a study treatment different from or in addition to the assigned treatment
- Lost to follow-up:
  - a subject who cannot be followed for some outcome or observation of interest
- In terms of its potential effect on the *validity* of the study, loss to follow-up is a more critical issue than dropping out of or into treatment groups
- Analyze on basis of the intention-to-treat principle

Adjusting the Sample Size

- For loss to follow-up, can scale up N by the proportion of subjects one anticipates may be lost to follow-up
- For drop-in and drop-out of treatment groups, make an allowance for the estimated (guessed) proportions of subjects dropping in and out and then adjust Δ
- Example: Clinical trial treating alcohol dependence
  - Outcome measure: Percent days abstinent (PDA)
  - Effective treatment assumed to increase mean PDA by 10 percentage points
  - Allowing for 25% drop-out of treatment, net effect is a mean increase of 7.5 percentage points

Adjusting the Sample Size: Example

- A clinical trial was planned to investigate reducing the risk of new cardiovascular events in subjects with history of CVD and periodontal disease
- Treatment groups: Study-supplied periodontal therapy versus usual care from own dentist
- Outcome measure: CVD event rate
- Effective treatment assumed to reduce rate by 25%
- Assume event rate of 6.5% p.a. in usual care group
- So assumed rate in active group is 4.875% p.a.

Adjusting the Sample Size: Example cont.

- Assume 10% in active care group drop out of treatment and 5% in usual care group drop in
- Net event rate in usual care group:
  - 95% have rate 6.5%; 5% have rate 4.875%
  - net rate: $0.95 \times 6.5\% + 0.05 \times 4.875\% = 6.42\%$
- Net event rate in active treatment group:
  - 90% have rate 4.875%; 10% have rate 6.5%
  - net rate: $0.9 \times 4.875\% + 0.1 \times 6.5\% = 5.04\%$
- Use the net rates in the power / sample size calculations

Correlation Coefficient / Linear Regression

- Let ρ be the correlation between X and Y and r be the sample correlation based on n pairs of observations
- Fisher's transform is approximately normally distributed:

$$Z_r = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right) \sim N\left(\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{N-3}\right)$$

- For the test $H_0: \rho = \rho_0$ versus $H_A: \rho \neq \rho_0$ to have power $1-\beta$ for the specific alternative $\rho = \rho_1$, we need a sample of size

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(Z_{\rho_1} - Z_{\rho_0})^2} + 3$$

Correlation Coefficient / Linear Regression

- For a simple linear regression model

$$Y = \beta_0 + \beta_1 X + \epsilon$$

we have

$$\widehat{\beta}_1 = \frac{s_Y}{s_X} r$$

or

$$r = \frac{s_X}{s_Y} \widehat{\beta}_1$$

where s_X and s_Y are the sample standard deviations of X and Y

Linear Regression Example

- Suppose we want to find the appropriate sample size to have 90% power to detect $\beta_1 = 0.5$ and from previous studies we have estimates $s_X = 2$ and $s_Y = 10$
- If $\widehat{\beta}_1 = 0.5$ then

$$r = \frac{s_X}{s_Y}\widehat{\beta}_1 = \frac{2}{10} \cdot 0.5 = 0.1$$

- We should have the same power to test

$$H_A : \beta_1 = 0.5 \;\text{ against }\; H_0 : \beta_1 = 0$$

as to test

$$H_A : \rho = 0.1 \;\text{ against }\; H_0 : \rho = 0$$

Linear Regression Example

- Here
$$Z_0 = \frac{1}{2}\log\left(\frac{1+0}{1-0}\right) = 0$$
and
$$Z_1 = \frac{1}{2}\log\left(\frac{1+0.1}{1-0.1}\right) = 0.1003$$

- The sample size n to give us power $1-\beta$ for testing $\rho = 0.1$ versus $\rho = 0$ is
$$\begin{aligned} n &= \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(Z_{\rho_1} - Z_{\rho_0})^2} + 3 \\ &= \frac{(1.96 + 1.28)^2}{(0.1003 - 0)^2} + 3 \\ &= 1046 \end{aligned}$$

Adjusting for Covariates in Regression Models

Reference: Hsieh, Bloch, Larsen (1998) A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine* 17:1623-1634.

- Suppose that our primary exposure of interest is X_1 and we want to adjust for covariates, $X_2, X_3, X_4, \ldots$
- Calculate R^2 from a regression model of X_1 (not Y) as a function of the other covariates
- Adjust the sample size using a variance inflation factor (VIF):

$$\text{VIF} = \frac{1}{1 - R^2}$$

  and use sample size $n' = \text{VIF} \cdot n$ where n is the sample size for a simple linear regression of Y on X_1

Example Adjusting for Covariates

- Continuing the previous example, suppose we want to adjust for just one additional variable X_2, with our interest still in the sample size to give 90% power to detect $\beta_1 = 0.5$
- Suppose the correlation between X_1 and X_2 is $r = 0.3$
- Then

$$n' = \text{VIF} \cdot n = \frac{1}{1 - R^2} \cdot n = \frac{1}{1 - r^2} \cdot n$$

$$= \frac{1}{1 - 0.3^2} \cdot 1046 = 1149$$