# Power and Sample Size II
# Bios 662

Michael G. Hudgens, Ph.D.

mhudgens@bios.unc.edu

http://www.bios.unc.edu/∼mhudgens

2008-11-03 13:13

# Outline

- Two sample: Continuous
- Two sample: Binary
- Case-control studies
- Estimating power with simulations

# Two sample test: Continuous Outcome

- Hyps

$$H_0 : \mu_1 = \mu_2 \text{ vs } H_A : \mu_1 \neq \mu_2$$

- Assume homogeneity of variance, $\sigma^2$ known, normality/CLT
- Then

$$N = \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_1 - \mu_2)^2} = 2\left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\Delta}\right)^2$$

- Note there are $N$ observations in each group such that total sample size is $2N$

# Two sample test: Continuous Outcome

- A drug company is comparing 2 drugs for lowering LDL-cholesterol
- Previous studies have found $\sigma^2 = 25^2 = 625$
- A difference of 15 mg/dl is assumed to be important
- For $\alpha = 0.05$ (2-sided) and $1 - \beta = .9$

$$N = \frac{2(625)(1.96 + 1.28)^2}{225} \approx 59$$

Thus 118 needed for study

## Two sample test: Continuous Outcome with $\sigma$ unknown

- What if the variance is not known?
- For $N_1 = N_2 = N$, can show

$$\frac{\overline{Y}_1 - \overline{Y}_2}{s_p\sqrt{\frac{2}{N}}} \sim t_{2N-2,\lambda}$$

where

$$\lambda = \Delta\sqrt{\frac{N}{2}}$$

# Two sample test: Continuous Outcome with $\sigma$ unknown

- R

```
# by hand
> 1-pt(qt(.975,116), 116, 15/25*sqrt(59/2))
[1] 0.8982732

> power.t.test(59, delta=15, sd=25)

     Two-sample t test power calculation

              n = 59
          delta = 15
             sd = 25
      sig.level = 0.05
          power = 0.8982732
    alternative = two.sided

 NOTE: n is number in *each* group
```

# Two sample test: Continuous Outcome with $\sigma$ unknown

- SAS

```
proc power;
   twosamplemeans
   meandiff   = 15
     ntotal = 118
     stddev = 25
     power  = .;
run;

/* output
Computed Power
Power
0.898
*/
```

## Two sample test: Continuous Outcome with $\sigma$ unknown

- Given $\beta$, solve for $N$

$$1 - \beta = \Pr[T \geq t_{2N-2,0;1-\alpha/2}]$$

  where $T \sim t_{2N-2,\Delta\sqrt{N/2}}$

- E.g., suppose $\beta = .1$, $\Delta = .5$. Numerical search in R:

```
> N <- 50; 1-pt(qt(.975,2*N-2), 2*N-2, 1/2*sqrt(N/2))
[1] 0.6968888
> N <- 90; 1-pt(qt(.975,2*N-2), 2*N-2, 1/2*sqrt(N/2))
[1] 0.9155872
> N <- 86; 1-pt(qt(.975,2*N-2), 2*N-2, 1/2*sqrt(N/2))
[1] 0.9032299
> N <- 85; 1-pt(qt(.975,2*N-2), 2*N-2, 1/2*sqrt(N/2))
[1] 0.899894
```

# Two sample test: Continuous Outcome with $\sigma$ unknown

- R

```
> power.t.test(power=.9, delta=.5)

     Two-sample t test power calculation

              n = 85.03129
          delta = 0.5
             sd = 1
      sig.level = 0.05
          power = 0.9
    alternative = two.sided
```

# Two sample test: Continuous Outcome with $\sigma$ unknown

- SAS

```
proc power;
   twosamplemeans
   meandiff   = 15
     ntotal = .
     stddev = 30
     power  = .9;
run;

/* output

Computed N Total

Actual         N
 Power    Total

 0.903      172
*/
```

# Two sample test: Binary Outcome

- Hyps

$$H_0 : \pi_1 = \pi_2 \text{ vs } H_A : \pi_1 \neq \pi_2$$

- Then

$$N \approx \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{(\pi_1 - \pi_2)^2}$$

where

$$\sigma^2 = \frac{1}{2}\{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)\}$$

see page 161 of text

- Again there are $N$ observations in each group such that total sample size is $2N$

# Two sample test: Binary Outcome

- Suppose $\pi_1 = .2727$, $\pi_2 = .2$, $\alpha = 0.05$ (two-sided), $1 - \beta = .9$. Then

$$N \approx \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{(\pi_1 - \pi_2)^2} = 712$$

- SAS

```
proc power;
   twosamplefreq
     refp = .2
     pdiff   = .0727
     ntotal = .
     power  = .9;
run;
```

# Two sample test: Binary Outcome

The POWER Procedure
Pearson Chi-square Test for Two Proportions

Fixed Scenario Elements

| | |
|---|---|
| Distribution | Asymptotic normal |
| Method | Normal approximation |
| Reference (Group 1) Proportion | 0.2 |
| Proportion Difference | 0.0727 |
| Nominal Power | 0.9 |
| Number of Sides | 2 |
| Null Proportion Difference | 0 |
| Alpha | 0.05 |
| Group 1 Weight | 1 |
| Group 2 Weight | 1 |

Computed N Total

| Actual Power | N Total |
|---|---|
| 0.900 | 1432 |

# Two sample test: Binary Outcome

- Why the difference? SAS is using a different approximation, which we now derive (cf Fleiss 1981)
- For $N_1 = N_2 = N$, Pearson's chi-square test statistic is equivalent to

$$Z = \frac{p_2 - p_1}{\sqrt{2\bar{p}\bar{q}/N}}$$

  where $\bar{p} = (p_1 + p_2)/2$, $\bar{q} = 1 - \bar{p}$
- WLOG, consider alternative $\pi_2 - \pi_1 = \delta_A > 0$.

# Two sample test: Binary Outcome

- Power to detect $\delta_A$ of two-sided test

  $\Pr[Z > z_{1-\alpha/2}|\delta_A]+\Pr[Z < z_{\alpha/2}|\delta_A] \approx \Pr[Z > z_{1-\alpha/2}|\delta_A]$

- Need to know distribution of $Z$ under $H_A$

$$E(p_2 - p_1) = \delta_A$$

$$Var(p_2 - p_1) = \frac{\pi_2(1-\pi_2)}{N} + \frac{\pi_1(1-\pi_1)}{N}$$

# Two sample test: Binary Outcome

$$1-\beta = \Pr\left[\frac{p_2-p_1}{\sqrt{2\bar{p}\bar{q}/N}} > z_{1-\alpha/2} \mid \delta_A\right]$$

$$= \Pr\left[p_2-p_1 > z_{1-\alpha/2}\sqrt{2\bar{p}\bar{q}/N} \mid \delta_A\right]$$

$$= \Pr\left[\frac{(p_2-p_1)-\delta_A}{\sqrt{Var(p_2-p_1)}} > \frac{z_{1-\alpha/2}\sqrt{2\bar{p}\bar{q}/N}-\delta_A}{\sqrt{Var(p_2-p_1)}} \mid \delta_A\right]$$

# Two sample test: Binary Outcome

- Implying

$$-z_{1-\beta} = \frac{z_{1-\alpha/2}\sqrt{2\bar{p}\bar{q}/N} - \delta_A}{\sqrt{Var(p_2 - p_1)}}$$

- Using $\bar{p}\bar{q} \approx \bar{\pi}(1-\bar{\pi})$ where $\bar{\pi} = (\pi_1 + \pi_2)/2$ yields

$$z_{1-\beta}\sqrt{Var(p_2 - p_1)} + z_{1-\alpha/2}\sqrt{2\bar{\pi}(1-\bar{\pi})/N} = \delta_A$$

# Two sample test: Binary Outcome

- Therefore

$$\frac{z_{1-\beta}\sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}+z_{1-\alpha/2}\sqrt{2\bar{\pi}(1-\bar{\pi})}}{\delta_A}=\sqrt{N}$$

- Thus, sample size required per arm to detect $\delta_A$ with $1-\beta$ power is

$$\frac{\{z_{1-\beta}\sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}+z_{1-\alpha/2}\sqrt{2\bar{\pi}(1-\bar{\pi})}\}^2}{\delta_A^2}$$

# Two sample test: Binary Outcome

- In R by hand

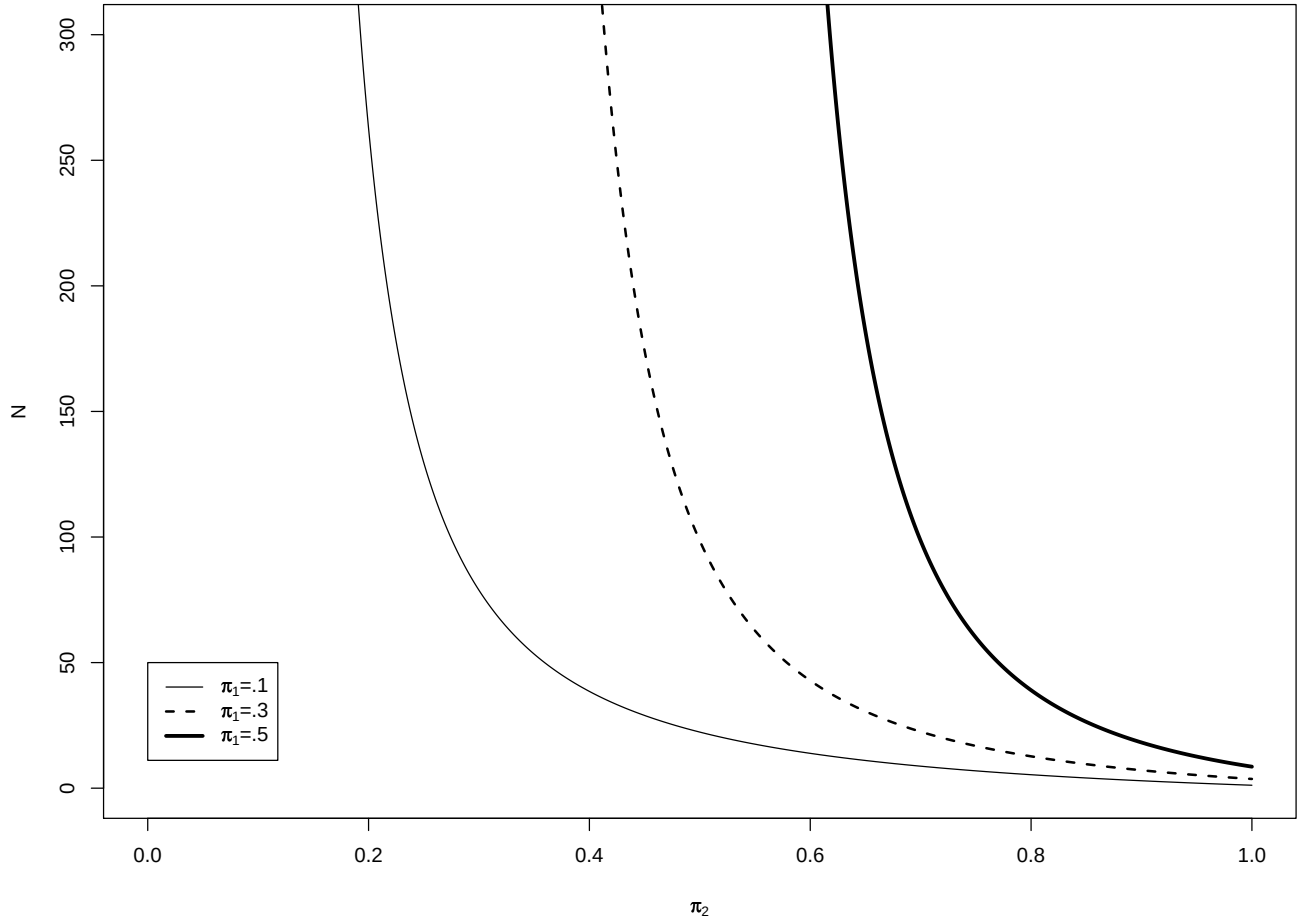
```
# sample size formula for comparing two
# binomial proportions based on fleiss (second edition) page 41

ss_fleiss <- function(pi1,pi2,alpha,power){
        q1 <- 1-pi1
        q2 <- 1-pi2
        pbar <- (pi1+pi2)/2
        qbar <- 1-pbar
        num <- qnorm(1-alpha/2)*sqrt(2*pbar*qbar)+qnorm(power)*sqrt(pi1*q1+pi2*q2)
        den <- (pi2-pi1)
        (num/den)^2
        }

> ss_fleiss(.2,.2727,.05,.9)
[1] 715.5618
```

# Graphical Summary

Sample size (per arm) for comparing $\pi_1$ with $\pi_2$ with $\alpha = .05$ (one-sided) and 90% power



# Case-Control: Binary Exposure

- Hyps

$$H_0 : OR = 1 \text{ vs } H_A : OR \neq 1$$

$$OR = \frac{odds(dis+|exp+)}{odds(dis+|exp-)}$$

- Recall

| | Disease | No disease |
|---|---|---|
| Exposed | $\pi_{11}$ | $\pi_{12}$ |
| Unexposed | $\pi_{21}$ | $\pi_{22}$ |

# Case-Control: Binary Exposure

$$OR = \frac{odds(dis+|exp+)}{odds(dis+|exp-)}$$

$$= \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}}$$

$$= \frac{\pi_{11}/\pi_{21}}{\pi_{12}/\pi_{22}}$$

$$= \frac{odds(exp+|dis+)}{odds(exp+|dis-)}$$

# Case-Control: Binary Exposure

- Hyps

$$H_0 : OR = 1 \text{ vs } H_A : OR \neq 1$$

- By previous slide

$$OR = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$$

  where $\pi_1 = \Pr(exp+|case)$, $\pi_2 = \Pr(exp+|control)$

- For specified $OR$ and $\pi_2$ we can determine $\pi_1$

$$\pi_1 = \frac{\pi_2 OR}{1+\pi_2(OR-1)}$$

# Case-Control: Binary Exposure

- Example:

  Cases: Neural tube defect babies

  Controls: Normal babies

  Exposure: Self reported dieting to lose weight during first trimester

- It is estimated that 20% of women will diet to lose weight during pregnancy; $\pi_2 = .2$
- The investigator wants to detect $OR = 1.5$

# Case-Control: Binary Exposure

- Solve for $\pi_1$

$$\pi_1 = \frac{.2(1.5)}{1 + .2(.5)} = 0.2727...$$

- With $\pi_1 = .2727$, $\pi_2 = .2$, $\alpha = 0.05$ (two-sided), $1-\beta = .9$, two sample binary example above yields $N = 712$ cases and $N = 712$ controls

# Case-Control Sample Size

- Cases often harder to obtain than controls
- How many controls per case?
- Continuous exposure model
- Discrete exposure model

# Case-Control: Continuous Exposure

- Ury (Biometrics 1975)
- Cases

$$Y_{1i} = \mu_i + \delta + \epsilon_{1i}; i = 1, \ldots, N$$

- Controls ($k$ for each case)

$$Y_{2ij} = \mu_i + \epsilon_{2ij}; j = 1, \ldots, k$$

- Assume $\epsilon_{1i}, \epsilon_{2ij}$ iid

$$E(\epsilon_{1i}) = E(\epsilon_{2ij}) = 0$$

$$V(\epsilon_{1i}) = V(\epsilon_{2ij}) = \sigma^2$$

# Case-Control: Continuous Exposure

- Let

$$\bar{Y}_{2i} = \frac{1}{k}\sum_{j=1}^{k} Y_{2ij}$$

- Then a consistent and unbiased estimator of exposure effect is

$$\hat{\delta}_k = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - \bar{Y}_{2i}) \equiv \bar{Y}_1 - \bar{Y}_2$$

# Case-Control: Continuous Exposure

- By independence and homogeneity of variance assumptions

$$V(\bar{Y}_1) = \frac{\sigma^2}{N} \text{ and } V(\bar{Y}_2) = \frac{\sigma^2}{kN}$$

- Therefore

$$V(\hat{\delta}_k) = \frac{\sigma^2}{N}\left(\frac{k+1}{k}\right)$$

- For $k = 1$,

$$V(\hat{\delta}_1) = \frac{2\sigma^2}{N}$$

# Case-Control: Continuous Exposure

- Relative efficiency

$$eff(\hat{\delta}_1, \hat{\delta}_k) = \frac{V(\hat{\delta}_k)}{V(\hat{\delta}_1)} = \frac{k+1}{2k} \rightarrow \frac{1}{2} \text{ as } k \rightarrow \infty$$

| $k$ | $eff(\hat{\delta}_1, \hat{\delta}_k)$ |
|---|---|
| 1 | 1.00 |
| 2 | 0.75 |
| 3 | 0.67 |
| 4 | 0.63 |
| 5 | 0.60 |
| 10 | 0.55 |
| $\infty$ | 0.50 |

# Case-Control: Continuous Exposure

- Assuming $N$ large or $\epsilon_{1i}, \epsilon_{2ij} \sim N(0, \sigma^2)$
- Under $H_0 : \delta = 0$

$$Z = \frac{\hat{\delta}_k}{\sqrt{V(\hat{\delta}_k)}} \sim N(0, 1)$$

- Under $H_A : \delta = \delta_A > 0$,

$$1 - \beta = \Pr\left[\frac{\hat{\delta}_k - \delta_A}{\sqrt{V(\hat{\delta}_k)}} > z_{1-\alpha/2} - \frac{\delta_A}{\sqrt{V(\hat{\delta}_k)}}\right]$$

# Case-Control: Continuous Exposure

- Implying

$$-z_{1-\beta} = z_{1-\alpha/2} - \frac{\delta_A}{\sqrt{V(\hat{\delta}_k)}}$$

$$(z_{1-\alpha/2} + z_{1-\beta})^2 = \frac{\delta_A^2}{V(\hat{\delta}_k)} = \delta_A^2 \frac{Nk}{\sigma^2(k+1)}$$

$$N = \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta_A^2} \frac{k+1}{2k}$$

# Case-Control: Continuous Exposure

- So, for two sample problem, compute usual sample size $N$ per arm assuming equal sample size per arm
- Multiply $N$ by $(k+1)/(2k)$ to get number of cases
- Multiply $N$ by $(k+1)/2$ to get number of controls

# Case-Control: Discrete Exposure

- The same relative efficiency result holds (Ury Biometrics 1975)
- Here comparing 1:1 vs $k$:1 controls:cases using generalization of McNemar/MH
- Same sample size computation; Cf Note 17.2 text

# Case-Control: Discrete Exposure

- Suppose with one control per case, we calculate 712 cases and 712 controls are needed to achieve a particular $\alpha$ and $\beta$
- Then with 2 controls per case, we need 712*3/4 = 534 cases and 1068 controls

# Outline

- Two sample: Continuous
- Two sample: Binary
- Case-control studies
- Estimating power with simulations

# Power and Sample Size

- Determination of power/SS important for many reasons
- Under-power: miss scientifically meaningful differences
- Over-power: waste of resources
- How to compute power/SS in more complicated situations than those addressed in the notes/text? Eg, what is the power of the Kruskal-Wallis test for fixed sample size?

# Simulated Power

- One answer: simulation study

  1. Simulate a single data set under particular alternative
  2. Evaluate test statistic for simulated data set. Record whether reject $H_0$ or not.
  3. Repeat Steps 1 and 2 multiple times (e.g., 1000)
  4. Compute proportion of simulated data sets where $H_0$ rejected. This is an estimate of power.

# Simulated Power

- To make sure your simulation is working correctly, check the following:
  - Simulate data sets under the null. Then the proportion of simulated data sets where $H_0$ rejected should not exceed the specified type I error rate $\alpha$
  - As you move away from $H_0$, the estimated power should increase towards 1

# Two sample test: Continuous Outcome with $\sigma$ unknown

- Recall

```
> power.t.test(59, delta=15, sd=25)

     Two-sample t test power calculation

              n = 59
          delta = 15
             sd = 25
      sig.level = 0.05
          power = 0.8982732
    alternative = two.sided

NOTE: n is number in *each* group
```

- Let's compare estimated power to this result

# Simulated power using R

```
n <- 59
sd <- 25
nsims <- 10000
mysim <- function(mdiff){
        rejects <- 0
        for (ii in 1:nsims){
                y1 <- rnorm(n,0,sd)
                y2 <- rnorm(n,mdiff,sd)
                tt <- t.test(y1,y2,var.equal=T)
                if (tt$p.value<0.05) rejects <- rejects + 1
                }
        print(paste("mdiff:",mdiff,", estimated power:",rejects/nsims))
        }
mysim(0)
mysim(10)
mysim(15)
mysim(20)
```

```
[1] "mdiff: 0 , estimated power: 0.0488"
[1] "mdiff: 10 , estimated power: 0.5843"
[1] "mdiff: 15 , estimated power: 0.8924"
[1] "mdiff: 20 , estimated power: 0.9876"
```

# Simulated power using SAS

```
%macro epower(mdiff=);

%let i=1;       %let n=59;      %let sd=25;     %let nsims=1000;

data;
  %do i = 1 %to &nsims;
    i=&i;
    do j= 1 to &n;
      y=rannor(0)*&sd; group=1; output;
    end;
    do j= 1 to &n;
      y=rannor(0)*&sd + &mdiff; group=2; output;
    end;
  %end;

ods output ttests=ttests;
proc ttest; class group; var y; by i; run;
data ttests; set ttests;
  if method="Pooled";
  reject=0; if Probt<0.05 then reject=1;

proc freq data=ttests; tables reject; run;
%mend;

%epower(mdiff=15);
```