JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Power and Sample Size: Issues in Study Design

John McGready
Johns Hopkins University

Lecture Topics

Re-visit concept of statistical power

Factors influencing power

Sample size determination when comparing two means

Sample size determination when comparing two proportions

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

Power and Its Influences

Consider the following results from a study done on 29 women, all 35–39 years old

Sample Data

	n	**Mean systolic BP**	**SD**
OC users	8	132.8	15.3
Non-OC users	21	127.4	18.2

Continued

Example

Of particular interest is whether OC use is associated with higher blood pressure

Statistically speaking we are interested in testing :

$H_o: \mu_1 = \mu_2$ (same as) $H_o: \mu_1 - \mu_2 = 0$

$H_o: \mu_1 \neq \mu_2$ $H_o: \mu_1 - \mu_2 \neq 0$

Continued

Here μ_1 represents (pop) mean BP for OC users, μ_2 (pop) mean BP for women not using OC

H_o: $\mu_1 = \mu_2$

H_o: $\mu_1 \neq \mu_2$

(same as)

H_o: $\mu_1 - \mu_2 = 0$

H_o: $\mu_1 - \mu_2 \neq 0$

Continued

Let's unveil the results of this test from our data:

	n	Sample **Mean Systolic BP**	Sample **SD**
OC users	8	132.8	15.3
Non-OC users	21	127.4	18.2

2-sample t-test $p = .46$

The sample mean difference in blood pressures is 132.8 – 127.4 = 5.4

This could be considered scientifically significant, however, the result is not statistically significant (or even close to it!) at the $\alpha = .05$ level

Continued

Suppose, as a researcher, you were concerned about detecting a population difference of this magnitude if it truly existed

This particular study of 29 women has low power to detect a difference of such magnitude

Recall, from lecture four of SR1:

DECISION		TRUTH: H_o	TRUTH: H_A
	Reject H_o	Type I Error α-level	Power
	Not Reject H_o		Type II Error β

Power is probability of rejecting H_o when H_a is true

Continued

Power is a measure of "doing the right thing" when H_a is true!

Higher power is better (the closer the power is to 1.0 or 100%)

We can calculate power for a given study if we specify a specific H_a

Example

This OC/Blood pressure study has power of .13 to detect a difference in blood pressure of 5.4 or more, if this difference truly exists in the population of women 35-39 years old!

	n	Sample **Mean Systolic BP**	Sample **SD**
OC users	8	132.8	15.3
Non-OC users	21	127.4	18.2

2-sample t-test p = .46

Continued

Example

When power is this low, it is difficult to determine whether there is no statistical difference in population means or we just could not detect it

Another Example

Suppose, instead of measuring SBP on each of the 29 women in the previous study, we instead classified them as “hypertensive” or “not hypertensive”

Four of the eight OC users were found to be “hypertensive” as were four of the 21 non-OC users

Continued

Our sample estimates:

- ★ $\hat{p}_1 = 4/8 = .50\ (50\%)$
- ★ $\hat{p}_2 = 4/21 = .19\ (19\%)$

This is a sample percentage difference of 31%! (RR = 2.6)

This could be a scientifically interesting population difference!

Continued

Another Example

The p-value from Fisher's Exact Test for testing:

$H_o: P_1 = P_2$ (same as) $H_o: P_1 - P_2 = 0$

$H_o: P_1 \neq P_2$ $H_o: P_1 - P_2 \neq 0$

p = .16, not significant at the $\alpha = .05$ level

Continued

Another Example

This study only has power of .24 to detect a difference of at least 31% in the population proportions of women with hypertension

We can not be sure whether our high p-value is because there is no population difference, or we could not detect such a population difference

Again, the higher the power of a study, the better!

When comparing two groups, we can calculate the power of a study if we specify a specific alternative hypothesis

Examples:

- ★ $\mu_1 - \mu_2 = 5$
- ★ $p_1 - p_2 = .25$

Continued

Notes on Power

The power of a study to detect a significant difference in two groups depends on the magnitude of the difference we want to detect:

- ★ *The smaller the difference of interest, the lower the power*

In order to INCREASE power for a study comparing two group means, we need to do the following:

- *Increase the sample size in each group*

  AND/OR

- *Change the estimates of μ_1 and μ_2 so that the difference between the two ($\mu_1 - \mu_2$) is bigger*

Continued

In order to INCREASE power for a study comparing two group means, we need to do the following:

- ★ *Increase the α-level of the hypothesis test (functionally speaking, make it "easier to reject")*

Continued

In order to INCREASE the power for a study comparing two proportions, we need to . . .

- *Increase the sample size in each group*

  AND/OR

- *Change the estimates of* P_1 *and* P_2 *so that the difference between the two* ($P_1 - P_2$) *is bigger*

Continued

In order to INCREASE power for a study comparing two proportions, we need to . . .

- ★ *Increase the α-level of the hypothesis test*

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

Practice Problems

1. Define *statistical power* and explain some possible implications of having low power in a statistical study.

Continued

2. Explain the relationship between power and sample size in a study (all other factors remaining constant)

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

Practice Problem Solutions

1. Define statistical power and explain some possible implications of having low power in a statistical study.
   - ★ *In the context of hypothesis testing, statistical power is the probability of rejecting the null hypothesis when the alternative is the truth*
   - ★ *In other words, the probability of "rejecting when we should" and the probability of finding a difference if it really exists*

Continued

1. Define statistical power and explain some possible implications of having low power in a statistical study.
   - ★ *If we have a study with low power and fail to reject the null hypothesis (fail to find a difference), it is difficult to ascertain whether the null hypothesis is true (there truly is no difference) or whether the alternative hypothesis is the truth (there is a difference) but the study could not detect this difference*

Continued

2. Explain the relationship between power and sample size in a study (all other factors remaining constant)
    - ★ *A simple relationship: As sample size increases across both groups, the power of the study increases*

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section B

Sample Size Calculation when Comparing Group Mean

Blood pressure and oral contraceptives

- *Suppose we used data from the example in Section A to motivate the following question:*
- *Is oral contraceptive use associated with higher blood pressure among individuals between the ages of 35–39?*

Recall, the data:

		Sample	Sample
	n	**Mean Systolic BP**	**SD**
O.C. Users	8	132.8	15.3
Non-O.C. Users	21	127.4	18.2
2-sample t-test p = .46			

Continued

We think this research has a potentially interesting association

We want to do a bigger study

- ★ *We want this larger study to have ample power to detect this association, should it really exist in the population*

Continued

What we want to do is determine sample sizes needed to detect about a 5mm increase in blood pressure in O.C. users with 80% power at significance level $\alpha = .05$

- ★ *Using pilot data, we estimate that the standard deviations are 15.3 and 18.2 in O.C. and non-O.C. users respectively*

Continued

Here we have a desired power in mind and want to find the sample sizes necessary to achieve a power of 80% to detect a population difference in blood pressure of five or more mmHg between the two groups

Continued

We can find the necessary sample size(s) of this study if we specify . . .

- ★ *α-level of test (.05)*
- ★ *Specific H_a (μ_1, μ_2 and hence $\mu_1 - \mu_2$ – usually represents the minimum difference of interest)*
- ★ *Estimates of σ_1 and σ_2*
- ★ *The power we desire (.80)*

Continued

How can we specify H_a and estimate SDs?

- ★ *Researcher knowledge—experience makes for good educated guesses*
- ★ *Make use of pilot study data!*

Continued

Fill in blanks from pilot study

- α *-level of test (.05)*
- *Specific* H_a *(* $\mu_1 = 132.8$, $\mu_2 = 127.4$*; and hence* $\mu_1 - \mu_2 = 5.4$*)*
- *Estimates of* σ_1 *(= 15.3)* σ_2 *(=18.2)*
- *The power we desire (.80)*

Continued

Given this information, we can use Stata to do the sample size calculation

"sampsi" command

- *Command syntax (items in blue are numbers to be supplied by researcher)*

sampsi μ_1 μ_2, alpha (α) power (power) sd1 (σ_1) sd2 (σ_2)

```
. sampsi 132.8 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.3)
```

```
Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            m1 =    132.8
            m2 =    127.4
           sd1 =     15.3
           sd2 =     18.3
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      154
            n2 =      154
```

Continued

```
. sampsi 132.8 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.3)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                  and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            m1 =    132.8
            m2 =    127.4
           sd1 =     15.3
           sd2 =     18.3
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      154
            n2 =      154
```

Continued

```
. sampsi 132.8 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.3)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                and m2 is the mean in population 2
Assumptions:

        alpha =   0.0500  (two-sided)
        power =   0.8000
           m1 =    132.8
           m2 =    127.4
          sd1 =     15.3
          sd2 =     18.3
        n2/n1 =     1.00

Estimated required sample sizes:

           n1 =      154
           n2 =      154
```

Continued

Our results from Stata suggest that in order to detect a difference in B.P. of 5.4 units (if it really exists in the population) with high (80%) certainty, we would need to enroll 154 O.C. users and 154 non-users

This assumed that we wanted equal numbers of women in each group!

Continued

```
. sampsi 132.8 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.3)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            m1 =    132.8
            m2 =    127.4
           sd1 =     15.3
           sd2 =     18.3
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      154
            n2 =      154
```

Continued

Suppose we estimate that the prevalence of O.C. use amongst women 35–39 years of age is 20%

- *We wanted this reflected in our group sizes*

If O.C. users are 20% of the population, non-OC users are 80%

- *There are four times as many non-users as there are users (4:1 ratio)*

Continued

We can specify a ratio of group sizes in Stata

- *Again, using "sampsi" command*

sampsi μ_1 μ_2, alpha(α) power(power) sd1(σ_1) sd2(σ_2) ratio(n_2/n_1)

Continued

Pilot Study/Stata Results

```
. sampsi 132.8 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.3) ratio(4)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                and m2 is the mean in population 2
Assumptions:

        alpha =   0.0500  (two-sided)
        power =   0.8000
           m1 =    132.8
           m2 =    127.4
          sd1 =     15.3
          sd2 =     18.3
        n2/n1 =     4.00

Estimated required sample sizes:

           n1 =       86
           n2 =      344
```

Sample Size for Comparing Two Means

Graphs

- ★ *Motulsky book, p. 198*
- ★ *Altman book, p. 456-457*

Computer

- ★ *Stata "sampsi" command*

Formula

- ★ *Campbell book, p.157*
- ★ *Daly book, p.426*

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section B

Practice Problems

1. Let's employ Stata to experiment with power and sample size calculations:
   - ★ *For the O.C.–B.P. experiment, suppose researchers were interested in looking at finer differences in B.P. between the O.C.–users and non-users.*
   - ★ *Suppose a pilot study found the estimated mean B.P. in O.C. users to be 130.1 units and 127.4 units in the non-users.*

Continued

1. This difference is considered scientifically interesting but was not found to be statistically significant in the pilot study.
   - *Recall that the estimated standard deviation for O.C. users is 15.3 units.*
   - *For non-O.C. users it is 18.3.*

Continued

1. What would the necessary group sizes be to conduct a study with $\alpha = .05$ and 80% power to detect a difference of this size?
   - ★ *Assume equal numbers of O.C. users and non-users*

Continued

2. What would the necessary group sample sizes be to do the same study as in question one, but with three times the number of non-O.C. users as compared to O.C. users?

   How does the total sample size (both groups together) compare in this scenario relative to the situation with equal group sizes?

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section B

Practice Problem Solutions

1. Let's employ Stata to experiment with power and sample size calculations:
   - *For the O.C.–B.P. experiment, suppose researchers were interested in looking at finer differences in B.P. between the O.C. users and non-users*
   - *Suppose a pilot study found the estimated mean B.P. in O.C. users to be 130.1 units and 127.4 units in the non-users*

Continued

1. This difference is considered scientifically interesting but was not found to be statistically significant in the pilot study
   - ★ *The S.D. for the O.C. users was 15.3*
   - ★ *The S.D. for the non-O.C. users was 18.3*

Continued

1. What would the necessary group sizes be to conduct a study with $\alpha = .05$ and 80% power to detect a difference of this size?
   - *Assume equal numbers of OC users and non-users*

Continued

```
. sampsi 130.1 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.3)

Estimated sample size for two -sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two -sided)
         power =   0.8000
            m1 =    130.1
            m2 =    127.4
           sd1 =     15.3
           sd2 =     18.3
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      613
            n2 =      613
```

Continued

2. What would the necessary group sample sizes be to do the same study as in question one, but with three times the number of non-O.C. users as compared to O.C. users?
   - ★ *How does the total sample size (both groups together) compare in this scenario relative to the situation with equal group sizes?*

Continued

Practice Problem Solutions

```
. sampsi 130.1 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.3) ratio(3)
Estimated sample size for two -sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                and m2 is the mean in population 2
Assumptions:

       alpha =   0.0500  (two -sided)
       power =   0.8000
          m1 =    130.1
          m2 =    127.4
         sd1 =     15.3
         sd2 =     18.3
       n2/n1 =     3.00

Estimated required sample sizes:

          n1 =      373
          n2 =     1119
```

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section C

Sample Size Determination for Comparing Two Proportions

Sample Size for Comparing Two Proportions

A randomized trial is being designed to determine if vitamin A supplementation can reduce the risk of breast cancer

- *The study will follow women between the ages of 45–65 for one year*
- *Women were randomized between vitamin A and placebo*

What sample sizes are recommended?

If, in fact, vitamin A is effective, you want a high likelihood of obtaining a statistically significant difference (i.e., reject H_o) between vitamin A and placebo

- *In other words, we want HIGH POWER to detect this association, should it exist*

We are interested in testing:

$$H_0 : P_1 = P_2$$
$$H_a : P_1 \neq P_2$$

(same as)

$$H_0 : P_2 - P_1 = 0$$
$$H_a : P_2 - P_1 \neq 0$$

Let's specify $\alpha = .05$

What sample sizes do we need to have power of .80 (80%)?

Necessary Information

We can find the necessary sample size(s) of this study if we specify the following:

- *α-level of test (.05)*
- *Specific* H_a *(best guess for population values of* P_1, P_2*, and hence* $P_2 - P_1$ *) – "guesstimate" of treatment effect*
- *The power we desire (.80)*

Design a study to have 80% power to detect a 50% (relative) reduction in risk of breast cancer with vitamin A using a (two-sided) test with significance level α-level = .05

The breast cancer rate in the controls can be assumed to be 150/100,000 per year

- *Estimated from other studies*

Continued

A 50% reduction:

$$\frac{150}{100{,}000} \times 0.5 = \frac{75}{100{,}000}$$

Continued

You would need about 34,000 individuals per group

Breast Cancer Sample Size Calculation in Stata

"sampsi" command

sampsi p_1 p_2, alpha(α) power(power)

Continued

Breast Cancer Sample Size Calculation in Stata

```
. sampsi .00075 .0015, alpha(.05) power(.8)

Estimated sample size for two-sample comparison of proportions

Test Ho: p1 = p2, where p1 is the proportion in population 1
                and p2 is the proportion in population 2
Assumptions:

        alpha =   0.0500  (two-sided)
        power =   0.8000
           p1 =   0.0008
           p2 =   0.0015
        n2/n1 =   1.00

Estimated required sample sizes:

           n1 =    33974
           n2 =    33974
```

Continued

Breast Cancer Sample Size Calculation in Stata

```
. sampsi .00075 .0015, alpha(.05) power(.8)

Estimated sample size for two-sample comparison of proportions

Test Ho: p1 = p2, where p1 is the proportion in population 1
                and p2 is the proportion in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            p1 =   0.0008
            p2 =   0.0015
         n2/n1 =   1.00

Estimated required sample sizes:

            n1 =    33974
            n2 =    33974
```

Continued

Breast Cancer Sample Size Calculation in Stata

```
. sampsi .00075 .0015, alpha(.05) power(.8)

Estimated sample size for two-sample comparison of proportions

Test Ho: p1 = p2, where p1 is the proportion in population 1
                and p2 is the proportion in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            p1 =   0.0008
            p2 =   0.0015
         n2/n1 =   1.00

Estimated required sample sizes:

            n1 =    33974
            n2 =    33974
```

Continued

Breast Cancer Sample Size Calculation in Stata

You would need about 34,000 individuals per group!

Why so many?

- *Difference between two hypothesized proportions is very small:* $P_2 - P_1 = .00075$

Continued

Breast Cancer Sample Size Calculation in Stata

We would expect about 50 cancer cases among the controls and 25 cancer cases among the vitamin A group

Expected number of cases in each group:

Placebo $$\frac{150}{100{,}000} \times 34{,}000 = 51$$

Vitamin A $$\frac{75}{100{,}000} \times 34{,}000 \approx 25$$

Breast Cancer—Vitamin A Example Revisited

Suppose you want 80% power to detect only a 20% (relative) reduction in risk associated with vitamin A

Continued

Breast Cancer—Vitamin A Example Revisited

$P_2 = .0015$

$P_1 = .0015 * .8 = .0012$

Note: $P_2 - P_1$, the treatment effect we are trying to detect, has gotten smaller

- *So N should get bigger! Here,* $P_2 - P_1 = .0003$

Breast Cancer Sample Size Calculation in Stata

```
. sampsi .0012  .0015, alpha(.05) power(.8)

Estimated sample size for two-sample comparison of proportions

Test Ho: p1 = p2, where p1 is the proportion in population 1
                    and p2 is the proportion in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            p1 =   0.0012
            p2 =   0.0015
         n2/n1 =   1.00

Estimated required sample sizes:

            n1 =    241769
            n2 =    241769
```

Breast Cancer—Vitamin A Example Revisited

You would need about 242,000 per group!

We would expect 360 cancer cases among the placebo group and 290 among vitamin A group

An Alternative Approach—Design a Longer Study

Proposal

- ★ *Five-year follow-up instead of one year*

$$P_1 \approx 5 \text{ X } .0012 = .006$$

$$P_2 \approx 5 \text{ X } .0015 = .0075$$

Continued

An Alternative Approach—Design a Longer Study

Need about 48,000 per group

- *Yields about 290 cases among vitamin A and 360 cases among placebo*

Issue

- *Loss to follow-up*

Two drugs for treatment of peptic ulcer compared (Familiari, et al., 1981)

- *The percentage of ulcers healed by pirenzepine (A) and trithiozine (B) was 76.7% and 58.1% based on 30 and 31 patients respectively (p = .17), 95% CI for difference in heal rates*
- *(A-B) was (-.04, .42)*

Source: Campbell, J., Machin (1993), Familiari, et al. (1981), Clinical Trial

Continued

Redesign a new trial

- *Use* $P_1 = .75$ *and* $P_2 = .60$
- *80% power*
- $\alpha = .05$

sampsi .75 .60, alpha (.05) power (.8)

N = 165 patients per group

Continued

Suppose you wanted two times as many people on trithiozone (“Drug B”) as compared to pirenzephine (“Drug A”)

- ★ *Here, the ratio of sample size for Group 2 to Group 1 is 2*

Can use “ratio” option in “sampsi” command

Continued

Example

```
. sampsi .75 .60, alpha(.05) power(.80) ratio(2)

Estimated sample size for two-sample comparison of proportions

Test Ho: p1 = p2, where p1 is the proportion in population 1
                    and p2 is the proportion in population 2
Assumptions:

        alpha =   0.0500  (two-sided)
        power =   0.8000
           p1 =   0.7500
           p2 =   0.6000
        n2/n1 =   2.00

Estimated required sample sizes:

           n1 =      126
           n2 =      252
```

Specify

In order to use the computer to compute sample size for a study comparing binary outcome (proportions) between two groups, you will need:

- ★ *Your best estimate of the true population proportions* P_1 *and* P_2
- ★ $\delta = P_2 - P_1$ *is the "treatment effect" OR "effect size"*
- ★ *The power you desire to detect this treatment effect*
- ★ α*-level of study*

Notes on Sample Per Group for Two Proportions

Graphs

- *Altman book, p. 455-459*

Computer

- *Stata "sampsi" command*

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section C

Practice Problems

Suppose a study was conducted to examine the relationship between Vitamin C and the common cold

- *Of a total of 20 subjects, 10 are randomized to take vitamin C for one month, and 10 to take a placebo*

Continued

Suppose a study was conducted to examine the relationship between Vitamin C and the common cold

- *At the end of the one-month period, subjects are asked detailed questions about the presence of cold symptoms during the month and then classified as having had a cold or not (one subject in the vitamin C group was classified as having had a cold, as were three subjects in the placebo group)*

Continued

1. What is the estimated risk difference of getting a cold for the Vitamin C group as compared with the placebo group?

2. What group sizes would be needed to do a study with $\alpha = .05$ and 80% power to detect the difference from (1)?

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section C

Practice Problem Solutions

Suppose a study was conducted to examine the relationship between Vitamin C and the common cold

- *Of a total of 20 subjects, 10 are randomized to take vitamin C for one month, and 10 to take a placebo*

Continued

Suppose a study was conducted to examine the relationship between Vitamin C and the common cold

- *At the end of the one-month period, subjects are asked detailed questions about the presence of cold symptoms during the month and then classified as having had a cold or not (one subject in the vitamin C group was classified as having had a cold, as were three subjects in the placebo group)*

Continued

1. What is the estimated risk difference of getting a cold for the Vitamin C group as compared with the placebo group?

$$\hat{p}_{vitC} = 1/10 = .10 \; ; \; \hat{p}_{Plac} = 3/10 = .30$$

$$\text{Risk Diff} = .30 - .10 = .20 \; (20\%)$$

Continued

2. What group sizes would be needed to do a study with $\alpha = .05$ and 80% power to detect the difference from (1)?

```
. sampsi .1  .3, alpha(.05) power(.8)

Estimated sample size for two-sample comparison of proportions

Test Ho: p1 = p2, where p1 is the proportion in population 1
                and p2 is the proportion in population 2
Assumptions:

        alpha =   0.0500  (two-sided)
        power =   0.8000
           p1 =   0.1000
           p2 =   0.3000
        n2/n1 =   1.00

Estimated required sample sizes:

           n1 =       72
           n2 =       72
```

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section D

Sample Size and Study Design Principles: A Brief Summary

When designing a study, there is a tradeoff between :

- *Power*
- *α-level*
- *Sample size*
- *Minimum detectable difference (specific Ha)*

Industry standard—80% power, $\alpha = .05$

Continued

What if sample size calculation yields group sizes that are too big (i.e., can not afford to do study) or are very difficult to recruit subjects for study?

- *Increase minimum difference of interest*
- *Increase α-level*
- *Decrease desired power*

Continued

Sample size calculations are an important part of study proposal

- ★ *Study funders want to know that the researcher can detect a relationship with a high degree of certainty (should it really exist)*

Continued

Even if you anticipate confounding factors, these approaches are the best you can do and are relatively easy

Accounting for confounders requires more information and sample size has to be done via computer simulation—consult a statistician!

Continued

When would you calculate the power of a study?

- ★ *Secondary data analysis*
- ★ *Data has already been collected, sample size is fixed*
- ★ *Pilot Study—to illustrate that low power may be a contributing factor to non-significant results and that a larger study may be appropriate*

Continued

What is this specific alternative hypothesis?

- *Power or sample size can only be calculated for a specific alternative hypothesis*
- *When comparing two groups this means estimating the true population means (proportions) for each group*

Continued

What is this specific alternative hypothesis?

- ★ *Therefore specifying a difference between the two groups*
- ★ *This difference is frequently called minimum detectable difference or effect size, referring to the minimum detectable difference with scientific interest*

Continued

Where does this specific alternative hypothesis come from?

- ★ *Hopefully, not the statistician!*
- ★ *As this is generally a quantity of scientific interest, it is best estimated by a knowledgeable researcher or pilot study data*
- ★ *This is perhaps the most difficult component of sample size calculations, as there is no magic rule or "industry standard"*

FYI—Using Stata to Compute Power

I promised you in part A of this lecture that I would eventually show you how to compute the power to detect difference in a study that has already been conducted

The “sampsi” command is still the command for this—we just need to feed it slightly different information for it to compute power

In order to calculate power for a study comparing two group means, we need the following:

- *Sample size for each group*
- *Estimated (population) means, μ_1 and μ_2 for each group—these values frame a specific alternative hypothesis (usually minimum difference of scientific interest)*

Continued

In order to calculate power for a study comparing two group means, we need the following:

- *Estimated (population) sd's,* σ_1 *and* σ_2 *for each group*
- α*-level of the hypothesis test*

Continued

The Blood Pressure/Oral Contraceptive Example

Sample Data

	n	Mean systolic BP	SD
OC users	8	132.8	15.3
Non-OC users	21	127.4	18.2

What is power of this study to detect a difference in average SBPS of 5.4 mmHg

Continued

Fill in information below with results form this study

- *Sample size for each group (* $n_{oc} = 8$*,* $n_{non\text{-}oc} = 21$*)*
- *Estimated (population) means,* $\mu_{oc} = 132.8$ *and* $\mu_{non\text{-}oc} = 127.4$
- *Estimated (population) sd's,* $\sigma_{oc} = 15.3$ *and* $\sigma_{non\text{-}oc} = 18.2$ *for each group*
- α*-level of the hypothesis test (.05)*

```
. sampsi 132.8 127.4,sd1(15.3) sd2(18.2) n1(8) n2

Estimated power for two-sample comparison of mear

Test Ho: m1 = m2, where m1 is the mean in populat
                    and m2 is the mean in populat
Assumptions:

         alpha =   0.0500  (two-sided)
            m1 =    132.8
            m2 =    127.4
           sd1 =     15.3
           sd2 =     18.2
sample size n1 =        8
            n2 =       21
         n2/n1 =     2.63

Estimated power:

         power =   0.1268
```

In order to calculate power for a study comparing two proportions, we need . . .

- *Sample size for each group*
- *Estimated (population) proportions,* p_1 *and* p_2 *for each group*
- *These values frame a specific alternative hypothesis (it usually is the minimum difference of scientific interest)*
- α*-level of the hypothesis test*

Continued

Hypertension/oral contraceptives example:

- ★ *= 4/8 = .50 (50%)*
- ★ *= 4/21 = .19 (19%)*

Continued

Fill in information below with results form this study

- *Sample size for each group (* n_{oc} *= 8,* $n_{non\text{-}oc}$ *= 21)*
- *Estimated (population) proportions,* p_{oc} *= .50 and* = *.19* $p_{non\text{-}oc}$
- α*-level of the hypothesis test (.05)*

Sampsi Command

```
. sampsi .5 .19,n1(8) n2(21) alpha(.05)

Estimated power for two-sample comparison of proport

Test Ho: p1 = p2, where p1 is the proportion in popu
1
                    and p2 is the proportion in popu
2
Assumptions:

         alpha =   0.0500  (two-sided)
            p1 =   0.5000
            p2 =   0.1900
sample size n1 =        8
            n2 =       21
         n2/n1 =     2.63

Estimated power:

         power =   0.2378
```