JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section B

Sample Size Calculations when Comparing Group Means

Example

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

2

Pilot Study

- Recall, the data:

	Sample Data		
	n	Mean SBP	SD of SBP
OC users	8	132.8	15.3
Non-OC users	21	127.4	18.2

3

Pilot Study

- 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
- 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

4

Pilot Study

- 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

5

Pilot Study

- We can find the necessary sample size(s) of this study if we specify the following:
  - α-level of test (.05)
  - Specific values for μ_1 and μ_2 (specific H_A) and hence $d = \mu_1 - \mu_2$: this usually represents the minimum scientific difference of interest
  - Estimates of σ_1 and σ_2
  - The desired power (.80)

6

Pilot Study

- How can we specify $d = \mu_1 - \mu_2$ and estimate population SDs?
  - Researcher knowledge—experience makes for good educated guesses
  - Make use of pilot study data!

7

Pilot Study

- Fill in blanks from pilot study
  - α -level of test (.05)
  - Specific H_A (μ_{OC} = 132.8, $\mu_{NO\ OC}$ = 127.4) and hence $d = \mu_1 - \mu_2$ = 5.4 mmHg
  - Estimates of σ_{OC} (= 15.3) and $\sigma_{NO\ OC}$ (= 18.2)
  - The power we desire (.80)

8

Pilot Study

- Given this information, we can use Stata to do the sample size calculation
- "sampsi" command
  - Command syntax (items in italics are numbers to be supplied by researcher)
  - sampsi μ_1 μ_2, alpha(α) power(*power*) sd1(σ*1*) sd2(σ2)

9

Stata Results

- Blood pressure/OC example

```
. sampsi 132.8 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.2)

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 =     18.2
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      153
            n2 =      153
```

10

Pilot Study/Stata Results

- Our results from Stata suggest that in order to detect a difference in blood pressure of 5.4 units (if it really exists in the population) with high (80%) certainty, we would need to enroll 153 OC users and 153 non-users
- This assumes that we want equal numbers of women in each group!

11

Stata Results

- Blood pressure/OC example

```
. sampsi 132.8 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.2)

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 =     18.2
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      153
            n2 =      153
```

12

Pilot Study/Stata Results

- Suppose we estimate that the prevalence of OC use amongst women 35–39 years of age is 20%
  - We want this reflected in our group sizes
- If OC 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)

13

Pilot Study/Stata Results

- We can specify a ratio of group sizes in Stata
  - Again, using "sampsi" command with "ratio" option
  - sampsi μ_1 μ_2, alpha(α) power(*power*) $sd_1(\sigma_1)$ $sd_2(\sigma_2)$ ratio(n_2/n_1)

14

Stata Resulss

- Blood pressure/OC example

```
. sampsi 132.8 127.4, alpha(.05) power(.8) sd1(15.3) sd2(18.2) 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 =     18.2
         n2/n1 =     4.00

Estimated required sample sizes:

            n1 =       86
            n2 =      344
```

15