


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

2



Section A

Power and Its Influences

Example

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

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

4

Example

- Of particular interest is whether OC use is associated with higher blood pressure
- Statistically speaking we are interested in testing:
 - $H_0: \mu_{OC} = \mu_{NO\ OC}$ $H_a: \mu_{OC} - \mu_{NO\ OC} = 0$
 - $H_0: \mu_{OC} \neq \mu_{NO\ OC}$ $H_a: \mu_{OC} - \mu_{NO\ OC} \neq 0$
- Here μ_{OC} represents (population) mean SBP for OC users, $\mu_{NO\ OC}$ (population) mean BP for women not using OC

5

Example

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

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

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

6

Power and Sample Size: Issues in Study Design: John McGready

Example

- 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
- 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

7

Power

- Recall, from lecture four of SR1:

		TRUTH	
		H_0	H_A
Reject H_0	Type I error alpha-level	Power 1-beta	
		Type II error beta	
Not Reject H_0			

8

Power

- Recall, from lecture four of SR1:

		TRUTH	
		H_0	H_A
Reject H_0	Type I error alpha-level	Power 1-beta	
		Type II error beta	
Not Reject H_0			

9

Power

- Recall, from lecture four of SR1:

		TRUTH	
		H_0	H_A
Reject H_0	Type I error alpha-level	Power 1-beta	
		Type II error beta	
Not Reject H_0			

10

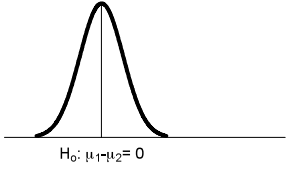
Power

- 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
- 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!
- 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

11

Power

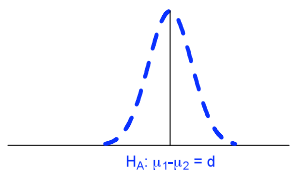
- Recall, the sampling behavior of $\bar{x}_1 - \bar{x}_2$ is normally distributed (large samples) with this sampling distributed centered at true mean difference
 - If H_0 truth, then curve is centered at 0



12

Power

- Recall, the sampling behavior of $\bar{x}_1 - \bar{x}_2$ is normally distributed (large samples) with this sampling distributed centered at true mean difference
 - If H_A truth, then curve is centered at some value d , $d \neq 0$

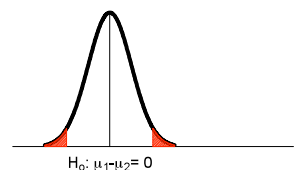


$H_A: \mu_1 - \mu_2 = d$

13

Power

- H_0 will be rejected (for $\alpha = .05$) if the sample result, $\bar{x}_1 - \bar{x}_2$, is more than 2 standard errors away from 0, either above or below

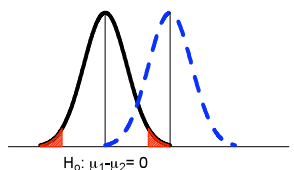


$H_0: \mu_1 - \mu_2 = 0$

14

Power

- H_0 will be rejected (for $\alpha = .05$) if the sample result, $\bar{x}_1 - \bar{x}_2$, is more than 2 standard errors away from 0, either above or below

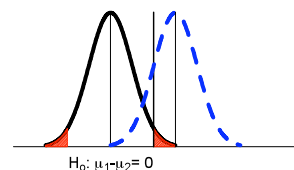


$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

15

Power

- H_0 will be rejected (for $\alpha = .05$) if the sample result, $\bar{x}_1 - \bar{x}_2$, is more than 2 standard errors away from 0, either above or below

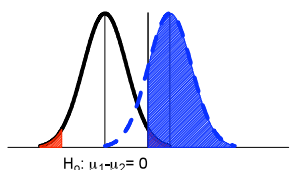


$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

16

Power

- H_0 will be rejected (for $\alpha = .05$) if the sample result, $\bar{x}_1 - \bar{x}_2$, is more than 2 standard errors away from 0, either above or below

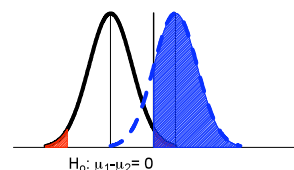


$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

17

What Influences Power?

- In order to increase power for a study comparing two group means, we need to do the following:
 - Change the estimates of μ_1 and μ_2 so that the difference between the two ($\mu_1 - \mu_2$) is bigger



$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

18

What Influences Power?

- In order to increase power for a study comparing two group means, we need to do the following:
 - Change the estimates of μ_1 and μ_2 so that the difference between the two ($\mu_1 - \mu_2$) is bigger

$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

19

What Influences 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

$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

20

What Influences 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

$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

21

What Influences Power?

- 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") here, with $\alpha = .10$:

$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

22

What Influences Power?

- 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") here, with $\alpha = .10$:

$H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 = d$

23

Power and Studies

- Power can be computed after a study is completed
 - Can only be computed for specific $H_{A,S}$ (i.e., this study had XX% to detect a difference in population means of YY or greater)
 - Sometimes presented as an "excuse" for non statistically significant finding ("the lack of a statistically significant association between A and B could be because of low power [$< 15\%$] to detect a mean difference of YY or greater between")
 - Can also be presented to corroborate with a non-statistically significant result
- Many times, in study design, a required sample size is computed to actually achieve a certain preset power level to find a "clinically /scientifically" minimal important difference in means
 - In the next section we will show how to do this using Stata
 - "Industry standard" for power: 80% (or greater)

24