Lecture 13

Designing Studies To Have A Desired Power

Learning Objectives

■ In this set of lectures the relationship between sample sizes and precision will re-expressed through the window of study power

■ It is more common to design a study to have a certain level of power (80% or 90%), than for a desired margin of error, but the approach is analogous

■ In these lecture sets, power and it's influences will be explored, and some examples of designing a study to achieve a certain power level will be given

2

Section A

Power and Its Influences

Example 1: Studies With Low Power

■ 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 1

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

■ Statistically speaking we are interested in testing :

$H_o: \mu_{OC} = \mu_{NO\ OC}$ $\qquad$ $H_o: \mu_{OC} - \mu_{NO\ OC} = 0$

$H_A: \mu_{OC} \neq \mu_{NO\ OC}$ $\qquad$ $H_A: \mu_{OC} - \mu_{NO\ OC} \neq 0$

Here μ_{OC} represents (population) mean SBP for OC users, $\mu_{NO\ OC}$ (population) mean BP for women not using OC

5

Example 1

■ Study results:

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

6

Example

- The sample mean difference in blood pressures is $132.8 - 127.4 = 5.4$ mmHg

- The 95% CI for the population level mean difference is (-8.9 mmHg, 19.7 mmHg), and the p-value is 0.43

7

Example

- 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

8

Power

- Recall the table comparing underlying truth to the decision made via hypothesis testing:

TRUTH

	H_o	H_A
Reject H_o		
Not Reject H_o		

9

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%), but comes at a "cost"

10

Power

- When a study with low power finds a non-statistically significant result, it is hard to interpret this result

- When a study has high power, a non-statistically significant result can be interpreted more confidently as "no association"

11

Power

- 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!

12

2

Power

- Recall, the sampling behavior of estimates comparing two samples (mean difference, risk difference) or the log of estimates comparing two samples is normally distributed (large samples) with this sampling distributed centered at true difference. If H_o truth, then curve is centered at 0

- For designing a study to have a certain power, or estimating the power of a completed study, we have to be specific about the value of H_A

13

Power

- For doing a hypothesis test comparing two groups, the null and alternative are:

    H_o: no difference
    H_A: a difference

14

Power

- If H_o truth, then sampling distribution of the estimate is centered at 0
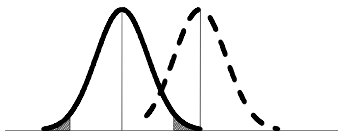
15

Power

- If H_A truth, then curve is centered at some value d, $d \neq 0$
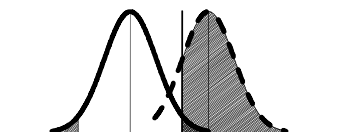
16

Power

- H_o will be rejected (for $\alpha = .05$) if the sample result is more than 2 standard errors away from 0, either above or below

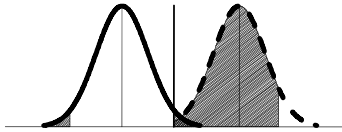

17

What Influences Power?

- In order to INCREASE power for a study comparing two populations, the researcher could
    Change the expected difference (specific H_A) to be larger

18

3

What Influences Power?
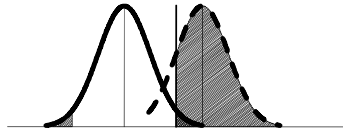
- In order to INCREASE power for a study comparing two populations, the researcher could
   Change the expected difference (specific H_A) to be larger

19

What Influences Power?

- In order to INCREASE power for a study comparing two populations, the researcher could
   Increase the sample size in each group

20

What Influences Power?

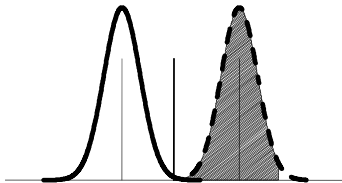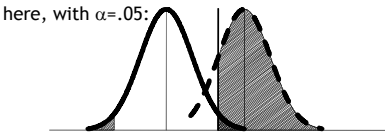- In order to INCREASE power for a study comparing two populations, the researcher could
   Increase the sample size in each group


21

What Influences Power?

- In order to INCREASE power for a study comparing two populations, the researcher could
   Increase the α-level of the hypothesis test (functionally speaking, make it "easier to reject")

here, with α=.05:

22

What Influences Power?

- In order to INCREASE power for a study comparing two populations, the researcher could
   Increase the α-level of the hypothesis test (functionally speaking, make it "easier to reject")

here, with α=.10:
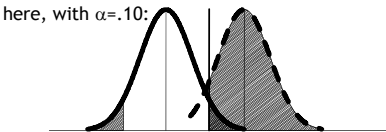
23

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 statically 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
   − "Industry standard" for power: 80% (or greater)

24

4

Power and Studies

- 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, proportions or incidence rates
  - "Industry standard" for power: 80% (or greater)

25

Summary

- The power of a study to detect a difference between populations on the appropriate measure of interest (difference in means, difference in proportions, relative risk, or incidence rate ratio) is a function of the size of the study samples, and the minimum detectable difference of interests

- When designing a study in advance, researchers need to incorporate these elements into the design while recognizing practical considerations such as budget and personnel

26

Section B

Sample Size Computations For Studies Comparing Two (or More) Means

Learning Objectives

- Upon completion of the lecture you will be able to
  - Describe the relationship between power and sample size with regards to the size of minimum detectable difference in means between two groups
  - Describe the relationship between power and sample size with regards to the standard deviation of individual values in the groups being compared
  - Understand the impact of designing studies to have equal versus on unequal sizes on the total sample size necessary to have a certain power

28

Example 1

- 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?

29

Example 1

- 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

30

Example 1

- 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 mean blood pressure in O.C. users with 80% power at significance level α = .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

31

Example 1

- 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 mean blood pressure of five or more mmHg between the two groups

32

Example 1

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

33

Example 1

- How can we specify $d= \mu_1$-μ_2 and estimate population SDs?
  - Researcher knowledge—experience makes for good educated guesses
  - Make use of pilot study data!

34

Example 1

- Fill in blanks from pilot study
  - α -level of test (0.05)
  - Specific H_A (μ_{OC}=132, $\mu_{NO\ OC}$ =127), and hence $d= \mu_1$-μ_2 =5 mmHg

  - Estimates of σ_{OC} (= 15.3) and $\sigma_{NO\ OC}$ (=18.2)
  - The power we desire (0.80)

35

Example 1

- Given this information, how can sample size be computed?
  - Statistical software such as Stata
  - Free online sample size calculators: a favorite from Dupont and Plummer at Vanderbilt University

  http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize

  - By hand (!?)

36

6

Example 1: Part 1

- For the first approach, let's assume we want equal numbers of women in each group

- Instead of taking one random sample from the clinical population of 35-39 year old women and then classifying each women as currently taking oral contraceptives (OCs) or not currently taking OCs, this approach would require taking two samples of women separately from those using OCs and those not currently using OCs

37

Example 1

- Based on using statistical software, we would need 178 women in each sample (for total of 356 women total) to have 80 % power to detect as large or larger than 5 mm Hg (in either direction)

- This corresponds to a margin of error of ± 3.6 mm Hg

38

Example 1

- Suppose we changed the minimum detectable difference to 4 mmHg. (for example $\mu_{OC}=132$, $\mu_{NO\ OC}=128$)

- Suppose we changed the minimum detectable difference to 6 mmHg. (for example $\mu_{OC}=132$, $\mu_{NO\ OC}=126$)

39

Example 1

- If a researcher was writing up a grant proposal, he/she may include a table like the following:

| | Sample Sizes Needed for Each Group for 80% Power | | |
| | Minimum Detectable Difference (mmHg) | | |
Estimated SDs	4	5	6
$sd_{OC}=14$, $sd_{NO\ OC}=17$	238	153	106
$sd_{OC}=15.3$, $sd_{NO\ OC}=18.2$	278	178	124
$sd_{OC}=16$, $sd_{NO\ OC}=19.5$	313	200	139

40

Example 1

- Suppose the funding agency reviewed the grant application, and asked for the same computations but for 90% power

| | Sample Sizes Needed for Each Group for 90% Power | | |
| | Minimum Detectable Difference (mmHg) | | |
Estimated SDs	4	5	6
$sd_{OC}=14$, $sd_{NO\ OC}=17$	319	204	142
$sd_{OC}=15.3$, $sd_{NO\ OC}=18.2$	372	238	166
$sd_{OC}=16$, $sd_{NO\ OC}=19.5$	418	268	186

41

Example 1: Part 2

- For the second approach, let's the sampling step includes taking a single representative sample of 35-39 year old women from this clinical population, and then classifying each as to her current OC usage

- In the original small study, 8 of the 29 women were currently using OCs: 28% of the sample. For purposes of the study let's use 30%. So we have to design a study that recognizes the unequal sample sizes.

42

Example 1

- Based on using statistical software, we would need 119 women in the OC group and 274 in the non-OC group (for total of 393 women total) to have 80 % power to detect a mean difference as large or larger than 5 mm Hg (in either direction)

- The total sample size (393) is larger than when the study was designed to have equal samples sizes (356). Why?

43

Example 1

- Suppose we changed the minimum detectable difference to 4 mmHg. (for example $\mu_{OC}=132$, $\mu_{NO\ OC} = 128$)

    $n_{OC}= 186$; $n_{NO\ OC} = 428$ ($n_{total} = 614$)

- Suppose we changed the minimum detectable difference to 6 mmHg. (for example $\mu_{OC}=132$, $\mu_{NO\ OC} =126$)

    $n_{OC}= 83$; $n_{NO\ OC} = 191$ ($n_{total} = 274$)

44

Example 2

- Suppose you are interested in designing a study to compared means between more than two groups. For example, you wish to compare the average length of stay for preventable diabetes hospitalizations across three insurance groups: government, private, and uninsured for diabetes patients in the state of Maryland in 2013. You plan to sample equal numbers from each of the groups.

- Based on data from another state, you have the following estimates:
    $\mu_{government} = 4.2$ days; $\mu_{private} = 3.1$ days; $\mu_{uninsured}= 2.5$ days;

The estimated standard deviations for the three groups are similar, at about 4 days.

45

Example 2

- How could a study be designed with 80% power to detect differences between the three groups

- One possibility: do the sample size computations for each unique two groups comparison, and take the maximum of the three computations

46

Example 2

- Sample size needed to 80% power

    Government vs. private: n = 208 in each group

    Government vs. uninsured: n= 87 in each group

    Private vs. no insurance: n= 698 in each group

47

Summary

- When designing a study to compare means from two or more populations, a researcher must have some estimate of the mean and standard deviation of the values in each population

- The sample size necessary to achieve a desired power to detect a minimum detectable difference is a function of the difference, the variability in the individual values in each group (standard deviation) and the desired power

48

8

Section C

Sample Size Computations For Studies Comparing Two (or More) Proportions or Incidence Rates

---

Learning Objectives

- Upon completion of the lecture you will be able to
  - Describe the relationship between power and sample size with regards to the size of minimum detectable difference in proportions or incidence rates between two groups
  - Understand the impact of designing studies to have equal versus unequal sizes on the total sample size necessary to have a certain power

50

Power for Comparing Two Proportions

- Same ideas as with comparing means, except that no standard deviation estimate is necessary (as the standard deviation of a proportion is a function of the proportion itself)

- We can find the necessary sample size(s) of this study if we specify .
  - α-level of test
  - Specific values for p_1 and p_2 (specific H_A) and hence d= p_1-p_2: usually represents the minimum scientific difference of interest)
  - The desired power

51

Example 1

- Two drugs for treatment of peptic ulcer compared (Familiari, et al., 1981)
  - The percentage of ulcers healed by pirenzepine (drug A) and trithiozine (drug B) was 77% and 58% based on 30 and 31 patients respectively (p-value = .17), 95% CI for difference in proportions healed $(p_{DRUG\ A} - p_{DRUG\ B})$ was(-.04, .42)
  - The power to detect a difference as large as the sample results with samples of size 30 and 31 respectively is only 25%

	Healed	Not Healed	Total
Drug A	23	7	30
Drug B	18	13	31

52

Example

- As a clinician, you find the sample results intriguing – want to do a larger study to better quantify the difference in proportions healed

- Redesign a new trial, using aforementioned study results to estimate population characteristics
  - Use $p_{DRUG\ A}$ = .77 and $p_{DRUG\ B}$ = .58 (RR =1.33)
  - 80% power
  - α = .05

53

Example 1

- As this is a randomized trial, to start let's assume equal sample sizes in the two groups

- Based on using statistical software, we would need 105 people in each sample (for total of 210 persons total) to have 80 % power to detect a difference in healing proportion as large or larger than 19%

- This corresponds to a margin of error of ±0.12 ($\pm12\%$)

54

Example 1

- Suppose we changed the minimum detectable difference to 10% mmHg. (for example $p_{DRUG\,A} = 0.77$, $p_{DRUG\,B} = 0.67$, RR =1.15)

    The sample size in each group is 335

- Suppose we changed the minimum detectable difference to 5 % mmHg. (for example $p_{DRUG\,A} = 0.77$, $p_{DRUG\,B} = 0.72$, RR=1.07)

    The sample size in each group is 1,232

55

Example

- Suppose you wanted to design a randomized clinical trial with two times as many people on trithiozone ("Drug B") as compared to pirenzephine ("Drug A"), to have 80% power to detect a difference of 19% (for example $p_{DRUG\,A} = .77$, $p_{DRUG\,B} = .58$)

    The study would require 80 people in the Drug A group and 160 in the Drug B group

56

Sample Size for Comparing Two Incidence Rates

- 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 will randomized between vitamin A and placebo

- What sample sizes are recommended?

57

Breast Cancer/Vitamin A Example

- Design a study to have 80% power to detect a 50% *relative reduction* in risk of breast cancer w/vitamin A

    (i.e. $IRR = \dfrac{IR_{VITA}}{IR_{PLACEBO}} = .50$)

using a (two-sided) test with significance level α-level = .05

- To get estimates of incidence rates of interest:

    - using other studies, the breast cancer rate in the controls can be assumed to be 150/100,000 per year

58

Breast Cancer/Vitamin A Example

- A 50% relative reduction: if

    $IRR = \dfrac{IR_{VITA}}{IRR_{PLACEBO}} = .50$ then $IR_{VIT\,A} = .50 \times IR_{PLACEBO}$

So, for this desired difference in the relative scale:

$IR_{VIT\,A} = \dfrac{150}{100,000}/\,year \times 0.5 = \dfrac{75}{100,000}/\,year$

59

Example 1

- As this is a randomized trial, to start let's assume equal sample sizes in the two groups

- Based on using statistical software, we would need 33,974 people in each sample (for total of 67,588 persons total) to have 80 % power to detect an incidence rate ratio of 0.5 or smaller

60

Breast Cancer Sample Size Calculation in Stata

- You would need about 34,000 individuals per group

- Why so many?
  - Difference between two hypothesized incidence rates is very small: 75 cases per 100,000 women (0.00075)

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

61

Breast Cancer/Vitamin A Example

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

A 20% relative reduction: if $IRR = \frac{IR_{VITA}}{IR_{PLACEBO}} = .80$ then $IR_{VIT\,A} = .80 \times IR_{PLACEBO}$

So, for this desired difference in the relative scale:

$$IRR_{VIT\,A} = \frac{150}{100,000} / year \times 0.8 = \frac{120}{100,000} / year$$

62

Example 1

- Again as this is a randomized trial, to start let's assume equal sample sizes in the two groups

- Based on using statistical software, we would need 241,889 people in each sample (for total of 483,798 persons total) to have 80 % power to detect an incidence rate ratio of 0.8 or smaller

63

Example 1

- You would need about 242,000 per group!

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

64

An Alternative Approach—Design a Longer Study

- Proposal
  - Five-year follow-up instead of one year
    Here:
    $IR_{VITA} \approx 5 \times .0012 = .006$ cases/5 years
    $IR_{PLACEBO} \approx 5 \times .0015 = .0075$/ 5 years

- Need about 48,000 per group
  - Yields about 290 cases among vitamin A and 360 cases among placebo

65

Summary

- When designing a study to compare proportions (or incidence rates) from two or more populations, a researcher must have some estimate of the expected proportion (or incidence rate) of the outcome in each population

- The sample size necessary to achieve a desired power to detect a minimum detectable difference in proportions (or incidence rate) is a function of the difference, and the desired power

66

Section D

Sample Size and Study Design Principles:
A Brief Summary

Designing Your Own Study

- When designing a study, there is a tradeoff between :
  - Power
  - α-level
  - Sample size
  - Minimum detectable difference (specific H_A)

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

68

Designing Your Own Study

- 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

69

Designing Your Own Study

- 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)

70

Designing Your Own Study

- 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

71

Designing Your Own Study

- 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 or incidence rates) for each group

72

Designing Your Own Study

- What is this specific alternative hypothesis?
  - This difference is frequently called minimum detectable difference or effect size, referring to the minimum detectable difference with scientific interest

73

Designing Your Own Study

- 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"

74

Planning for Loss to Follow-Up

- The sample size estimates given by the software do not account for potential drop out (loss to follow-up), or non-participation

- Frequently, researchers will add a buffer of 5-10% to the necessary sample sizes to achieve a desired level of power to allow for some dropout/non-participation

75

Section E

FYI if Interested (Optional)

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

77

Example

- Suppose we want to design a study with 80% power to detect a mean difference of at least 5 mmHg between the two group
  - will reject at α=.05 if $\left|\frac{\bar{x}_{oc} - \bar{x}_{no\,oc}}{SE(\bar{x}_{oc} - \bar{x}_{no\,oc})}\right| \geq 2$

  - as such, want $\Pr\left(\left|\frac{\bar{x}_{oc} - \bar{x}_{no\,oc}}{SE(\bar{x}_{oc} - \bar{x}_{no\,oc})}\right| \geq 2\right) = 0.80$ if $\mu_{OC}-\mu_{NO\,OC} \geq 5$ mmHg

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

78

13
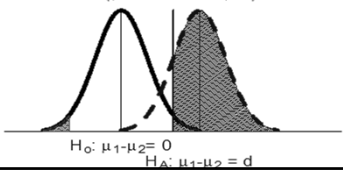
Example

- Consider $SE(\bar{x}_{OC} - \bar{x}_{NO\ OC})$

Using the estimates from the small study for population SDs:

$$SE(\bar{x}_{OC} - \bar{x}_{NO\ OC}) = \sqrt{\frac{\sigma^2_{OC}}{n_{OC}} + \frac{\sigma^2_{NO\ OC}}{n_{NO\ OC}}}$$

with $n_{OC} = n_{NO\ OC} = n$ this becomes:

$$SE(\bar{x}_{OC} - \bar{x}_{NO\ OC}) = \sqrt{\frac{\sigma^2_{OC}}{n} + \frac{\sigma^2_{NO\ OC}}{n}} = \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}$$

79

Example

- Suppose we want to design a study with 80% power to detect a mean difference of at least 5 mmHg between the two groups

- i.e. $\Pr\left(\left|\frac{\bar{x}_{OC} - \bar{x}_{NO\ OC}}{\sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}}\right| \geq 2\right) = 0.80$ if $\mu_{OC} - \mu_{NO\ OC} \geq 5$ mmHg

With some algebra:

$$\Pr\left(|\bar{x}_{OC} - \bar{x}_{NO\ OC}| \geq 2 \times \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}\right) = 0.80 \quad \text{if } \mu_{OC} - \mu_{NO\ OC} \geq 5.4 \text{mmHg}$$

80

Example

- Suppose we want to design a study with 80% power to detect a mean difference of at least 5 mmHg between the two groups

But if $\mu_{OC} - \mu_{NO\ OC} = 5$ mmHg, then assuming large n, $\bar{x}_{OC} - \bar{x}_{NO\ OC}$ is a

normally distributed process with mean $\mu_{OC} - \mu_{NO\ OC}$ and standard error

$$SE(\bar{x}_{OC} - \bar{x}_{NO\ OC}) = \sqrt{\frac{\sigma^2_{OC}}{n} + \frac{\sigma^2_{NO\ OC}}{n}} = \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}$$

81

Example

- Suppose we want to design a study with 80% power to detect a mean difference of at least 5 mmHg between the two groups

So $\Pr\left(|\bar{x}_{OC} - \bar{x}_{NO\ OC}| \geq 2 \times \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}\right) = 0.80$ if $\mu_{OC} - \mu_{NO\ OC} = 5$ mmHg

Becomes:

$$\Pr\left(\left|\frac{\bar{x}_{OC} - \bar{x}_{NO\ OC}}{\sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}}\right| \geq \frac{2 \times \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})} - (\mu_{OC} - \mu_{NO\ OC})}{\sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}}\right) = 0.8$$

82

Example

- Suppose we want to design a study with 80% power to detect a mean difference of at least 5 mmHg between the two groups

But on a standard normal curve, the value that cuts of 80% of the area to its right is 0.84.

So we need to solve: $\dfrac{2 \times \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})} - (\mu_{OC} - \mu_{NO\ OC})}{\sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}} = -0.84$

Some more beautiful algebra:

$$2 \times \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})} - (\mu_{OC} - \mu_{NO\ OC}) = -0.84 \times \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})}$$

83

Example

- Suppose we want to design a study with 80% power to detect a mean difference of at least 5 mmHg between the two groups

Some more beautiful algebra:

$$(2 + 0.84) \times \sqrt{\frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC})} = (\mu_{OC} - \mu_{NO\ OC})$$

squaring both sides:

$$(2 + 0.84)^2 \times \frac{1}{n}(\sigma^2_{OC} + \sigma^2_{NO\ OC}) = (\mu_{OC} - \mu_{NO\ OC})^2$$

$$\frac{(2 + 0.84)^2 \times (\sigma^2_{OC} + \sigma^2_{NO\ OC})}{(\mu_{OC} - \mu_{NO\ OC})^2} = n$$

84

14

Example

- Suppose we want to design a study with 80% power to detect a mean difference of at least 5 mmHg between the two groups

  Plugging on our info:
  $$\frac{(2 + 0.84)^2 \times (15.3^2 + 18.2^2)}{(5)^2} = n$$

  $$\frac{8.1 \times 565}{25} = n$$

  $$183 \approx n$$

85