Lecture 12

Designing Studies Such that Estimates Have A Desired Level of Precision (Margin of Error) - OPTIONAL

Section A

Precision and Sample Size: An Overview

Learning Objectives

- Explain and demonstrate empirically the relationship between sample size and precision of an estimate (ie: margin of error, confidence interval width)
- Explain that for continuous measures, the variability of the individual values (standard deviation) also impacts the precision of sample based estimates

3

Example 1

- Suppose a researcher takes a sample of 30 discharge records from 2012 for a large urban teaching hospital. The researcher wants to make a statement about length of stay at the facility.

Results: $\bar{x} = 30$ days; s = 7.5 days, n = 30

95% CI: $6.5 \pm 2.05 \times \frac{7.5}{\sqrt{30}}$

$\Rightarrow 6.5 \pm 2.05 \times 1.37$

$\Rightarrow 6.5 \pm 2.8$

$\Rightarrow$ (3.5 days, 9.1 days)

4

Example 2

Here is summary data on charges by sex based on a random sample of 500 Carotid Endarterectomy (CE) procedures performed in State of Maryland ,1995

	Males	Females
Mean Charges (US $)	6,615	7,088
SD (US $)	4,220	4908
N	271	229

5

Example 2

- Mean difference in charges, males to females

95% CI: $(7{,}088 - 6{,}615) \pm 2 \times \sqrt{\frac{(4{,}908)^2}{229} + \frac{(4{,}220)^2}{272}}$

$\Rightarrow 473 \pm 2 \times 413.40$

$\Rightarrow 473 \pm 826.80$

$\Rightarrow (-\$353.80, \$1{,}299.80)$

6

Example 3

- Researchers design a small pilot study to estimate the percentage of patients who have a minor reaction to a drug

30 subjects are enrolled; 9 have reaction ($\hat{p} = 0.30$)

95% CI: $0.30 \pm 2\sqrt{\frac{(.3)(.7)}{30}}$

$\Rightarrow 0.3 \pm 2 \times 0.084$

$\Rightarrow 0.3 \pm 0.168$

$\Rightarrow (0.13, 0.47)$

Example 4

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

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

$$\hat{p}_{drugA} = \frac{23}{30} = 0.77; \hat{p}_{drugA} = \frac{18}{31} = 0.58$$

Example 4

- Results $\hat{p}_{drugA} - \hat{p}_{drugA} = 0.19$

$0.19 \pm \sqrt{\frac{(.77)(.23)}{30} + \frac{(.58)(.42)}{31}}$

$\Rightarrow 0.19 \pm 2(0.116)$

$\Rightarrow 0.19 \pm .232$

$\Rightarrow (-0.042, 0.42)$

Example 5

- Suppose a small study is done to compare the efficacy of two smoking cessation programs. For program 1, 10 people enrolled and were following for a cumulative 259 days after quitting. Four persons resumed smoking in the follow-up period. For program 2, 11 people enrolled and were followed for a cumulative total of 267 days, with two persons resuming smoking in the follow-up period.

$$I\hat{RR} = \frac{\left(\frac{4}{259\text{days}}\right)}{\left(\frac{2}{267\text{days}}\right)} \approx 2.06$$

Example 5

- Results: $I\hat{RR} \approx 2.06; \ln(I\hat{RR}) = \ln(2.06) = 0.72$

95% CI for IRR: (0.36, 11.7)

Summary

- Small sample size(s) lead to larger standard errors and hence poorer precision of sample estimates (as estimates for underlying population level truths)
- We will see how to estimate the necessary sample size(s) to design a study to have a desired level of precision (margin of error, CI width)

Section B

Computing Sample Size to Achieve a Desired Level of Precision : Single Population Quantities

Learning Objectives

- Upon completion of this section you will be able to:
  - Create a table relating sample size to precision for an estimate of a single population quantity
  - Solve for the necessary sample size to get a desired level of precision

14

The Idea

- In order to justify the funding request for a larger study, a researcher needs to both demonstrate that the study:
  - Allows for the estimation of outcome(s) with a "good" margin of error
  - The study can performed given the requested budget

15

Inputs

- Designing a study, such the result(s) has a certain margin of error requires some speculation about what the study results will be, before the study is done!!??! Where can this information come from?

16

Example 1

- Results from the length of stay study with n= 30 subjects

$$\bar{x} = 6.3 \text{ days}; \ \ s = 7.5 \text{ days}$$

Margin of error:

- Margin of error estimate, as a generic function of n

17

Example 1

- Estimated margin of error for a study with n= 100
- Estimate margin of error for a study with n=250

18

Example 1

- You could easily make a table like the following:

	Standard Deviation		
n	6	7.5	9
100	1.20	1.50	1.80
200	0.85	1.06	1.27
300	0.69	0.87	1.04

19

Example 1

- You could easily solve for the sample size to get a desired margin of error. For example, suppose you want to be able to estimate the mean length of stay within ± 0.5 days

20

Example 2

- Recall the pilot study example from section A. Thirty participants were given a drug and followed to see who experienced a minor reaction; 9 subjects had the reaction.

$\hat{p} = 0.30$

- Margin of error estimate, as a generic function of n

21

Example 2

- Estimated margin of error for a study with n= 150

- Estimate margin of error for a study with n=300

22

Example 2

- You could easily make a table like the following:

	Expected Proportion		
n	0.25	0.3	0.35
100	0.087	0.092	0.095
200	0.061	0.065	0.067
300	0.050	0.053	0.055

23

Example 2

- You could easily solve for the sample size to get a desired margin of error. For example, suppose you want to be able to estimate the proportion of patients who have the reaction within ± 2.5 % (± 0.025)

24

Summary

- In order to compute margin of error for given sample size, you will need estimates of :
  - Standard deviation for a continuous measure
  - Proportion for a binary outcome
  - Incidence rate for a time-to event outcome (we did not show an example with an IR and the computations is a little trickier, but the principle is the same)
- The estimates of the aforementioned quantities came come from a small pilot study, secondary results from other research or educated guesswork

25

Section C

Computing Sample Size to Achieve a Desired Level of Precision : Population Comparison Measures

Learning Objectives

- Upon completion of this section you will be able to:
  - Have a general idea of how to create a table relating sample size(s) to precision for an estimate of a population comparison quantity (mean difference, difference in proportions)

27

The Idea

- In order to justify the funding request for a larger study, a researcher needs to both demonstrate that the study:
  - Allows for the estimation of outcome(s) with a “good” margin of error
  - The study can performed given the requested budget

28

Inputs

- Designing a study, such the result(s) has a certain margin of error requires some speculation about what the study results will be, before the study is done!!??! Where can this information come from?

29

Example 1

Here is summary data on charges by sex based on a random sample of 500 Carotid Endarterectomy (CE) procedures performed in State of Maryland ,1995

	Males	Females
Mean Charges (US $)	6,615	7,088
SD (US $)	4,220	4908
N	271	229

30

Example 1

- Results from the length of stay study with n= 500 subjects total

$$\bar{x}_{males} - \bar{x}_{females} = -473$$

- (95%) Margin of error estimate, as a generic function of n_1, n_2

$$2\hat{SE}(\bar{x}_{males} - \bar{x}_{females}) = 2\left(\sqrt{\frac{s^2_{males}}{n_{males}} + \frac{s^2_{females}}{n_{females}}}\right)$$

31

Example 1: Equal Sample Sizes

- Estimated margin of error for a study with n_{males}=$n_{females}$= 1,000

$$2\hat{SE}(\bar{x}_{males}\text{-}\bar{x}_{females}) = 2\left(\sqrt{\frac{4{,}220^2}{1{,}000} + \frac{4{,}908^2}{1{,}000}}\right) = 2(205) = \$410$$

- Estimate margin of error for a study with n_{males}=$n_{females}$= 1,800

$$2\hat{SE}(\bar{x}_{males}\text{-}\bar{x}_{females}) = 2\left(\sqrt{\frac{4{,}220^2}{1{,}800} + \frac{4{,}908^2}{1{,}800}}\right) = 2(153) = \$306$$

32

Example 1: Equal Sample Sizes

- The equation could also be solved for n_{males} (and hence $n_{females}$=n) given a desired margin of error, m

$$m = 2\left(\sqrt{\frac{4{,}220^2}{n} + \frac{4{,}908^2}{n}}\right) = 2\sqrt{\frac{4{,}220^2 + 4{,}908^2}{n}}$$

33

Example 1

- You could easily make a table like the following:

	Sample Sizes Needed for Each Group		
	Desired Margin of Error (US $)		
Estimated SDs	**200**	**250**	**300**
sd_{males}=$4,200 $sd_{females}$= $4,200			
sd_{males}=$4,220 $sd_{females}$= $4,900			
sd_{males}=$4,500 $sd_{females}$= $5,100			

34

Example 1

- This approach could easily be modified to allow for unequal sample sizes as well

35

Summary

- A very similar approach can be taken to estimate sample sizes necessary to have a desired margin of error for a risk difference (difference in proportions)
- For ratios, this process is a bit more difficult: however, when we get to the next section we'll show a more commonly employed method for estimating the necessary sample sizes via a metric called study power, which is related to margin of error

36