STA 291
Lecture 21

All confidence intervals we learned here is of the form

Point estimator $\pm$ error bound

Interchangeable wording:
Error bound = margin of error

- If everything else held unchanged;

increase confidence level → larger error bound

- If everything else held unchanged;

increase sample size n → smaller error bound

Planning on the sample size n

- Usually we first fix a confidence level, e.g. 95%.
- Then we would "trial and error" with different sample size *n* and see how small/large the error bound would be.

Example

- For 95% confidence intervals on a proportion p
- If n = 1500 → error bound = 0.02530 or 2.53%
- If n = 1000 → error bound=0.03099 or 3.10%

STA 291 - Lecture 21 4

- If n = 700 → error bound = 0.03704 or 3.7%
- If n = 500 → error bound = 0.0438 or 4.38%

Etc. etc. The formula I used is

STA 291 - Lecture 21 5

- Error bound = $1.96 \cdot \frac{\sqrt{0.5(1-0.5)}}{\sqrt{n}}$
- If there is no reliable information on p, we can use the conservative value: p= 0.5 (the answer is not very sensitive to the change in the value of p)

STA 291 - Lecture 21 6

Choice of sample size

- Margin of error = error bound = B

Choice of sample size

- In order to achieve a margin of error B, (with confidence level 95%), how large the sample size n must we get?
- For the confidence interval of population mean, mu, the formula is

Choice of Sample Size

$$\bar{X} \pm z \cdot \frac{s}{\sqrt{n}} = \bar{X} \pm B$$

- So far, we have calculated confidence intervals starting with *z*, *n* and s (plus, a possible t adjustment)
- These three numbers determine the error bound *B* of the confidence interval
- Now we reverse the equation:
  - We specify a desired error bound *B*
  - Given *z* and s, we can find the minimal sample size n needed for achieve this.

Choice of Sample Size

- From last page, we have

$$z \cdot \frac{s}{\sqrt{n}} = B$$

- Mathematically, we need to solve the above equation for *n*
- The result is

$$n = s^{2} \cdot \left(\frac{z}{B}\right)^{2}$$

- B must be in decimal form

Example

- About how large a sample would have been adequate if we merely needed to estimate the mean to within 0.5 unit, with 95% confidence?
- (assume $s = 5$ this may come from a pilot study)
- B=0.5, z=1.96
- Plug into the formula: $n = 5^{2} \cdot \left(\frac{1.96}{0.5}\right)^{2}$ =384.16

- In reality, s is usually replaced by *s* and

We need to replace *z* by *t* (with t-table).

For example, if the number 5 is actually *s*, not s then

$$n = 5^{2} \cdot \left(\frac{1.984}{0.5}\right)^{2} = 393.62$$

- I want to stress that these are somewhat approximate calculations, as they rely on the pilot information about either s or p, which may or may not be very reliable.
- But it is much better than no planning

Choice of sample size

- The most lazy way to do it is to guess a sample size *n* and
- Compute B, if B not small enough, then increase *n*;
- If B too small, then you may decrease *n*

- For the confidence interval for p

$$z \cdot \frac{\sqrt{p(1-p)}}{\sqrt{n}} = B$$

- Often, we need to put in a rough guess of p (called pilot value). Or, conservatively put p=0.5

- Suppose we want a 95%confidence error bound B=3% (margin of error + - 3%).

Suppose we do not have a pilot p value, so use p = 0.5

So, n = 0.5(1-0.5) [1.96/0.03]^2=1067.11

Example 1(from last lecture):

- Smokers try to quit smoking with Nicotine Patch or Zyban.
- Placebo 160 subjects, 30 quit
- Patch: 244 subjects, 52 quit
- Zyban: 244 subjects, 85 quit
- Zyban+patch: 245 subjects, 95 quit

- Find the 95% confidence intervals for p: the success rate/proportion

95% confidence intervals for p

- Placebo: [0.13, 0.25]
- Patch: [0.16, 0.26]
- Zyban: [0.29, 0.41]
- Zyban+patch: [0.33, 0.44]

Example 2

- To test a new, high-tech swimming gear, a swimmer is asked to swim twice a day, one with the new gear, one with the old.
- The difference in time is recorded:

Time(new) – time(old) = -0.08, -0.1, 0.02, …. -0.004. There were a total of 21 such differences.

Q: is there a difference?

- First: we recognize this is a problem with mean mu.
- And we compute the average X bar= -0.07
- SD = 0.02
- 90% confidence interval is:

Plug-in the values into formula

$$\bar{X} - ??\frac{0.02}{\sqrt{21}} \quad \text{and} \quad \bar{X} + ??\frac{0.02}{\sqrt{21}}$$

$$-0.07 - ??\frac{0.02}{\sqrt{21}} \quad \text{and} \quad -0.07 + ??\frac{0.02}{\sqrt{21}}$$

- What is the ?? Value.
- It would be 1.645 if we knew sigma, the population SD. But we do not, we only know the sample SD. So we need T-adjustment.
- Df= 21 -1 = 20
- ??=1.725

Example 3: Confidence Interval

- Example: Find and interpret the 95% confidence interval for the population mean, if the sample mean is 70 and the pop. standard deviation is 12, based on a sample of size

n = 100

First we compute $\frac{s}{\sqrt{n}}$ *=12/10= 1.2 ,*

1.96x 1.2=2.352

[70 – 2.352, 70 + 2.352] = [67.648, 72.352]

Example: Confidence Interval

- Now suppose the pop. standard deviation is unknown (often the case). Based on a sample of size *n = 100 , Suppose we also compute the s = 12.6 (in addition to sample mean = 70)*

First we compute $\frac{s}{\sqrt{n}}$ *=12.6/10= 1.26 ,*

From t-table 1.984 x 1.26 = 2.4998

[70 – 2.4998, 70 + 2.4998] = [67.5002, 72.4998]

Error Probability

- The error probability (a) is the probability that a confidence interval does **not** contain the population parameter -- (missing the target)
- For a 95% confidence interval, the error probability a=0.05
- a = 1 - confidence level or

confidence level = 1 – a

Different Confidence Levels

Confidence level	*Error a*	*a/2*	*z*
90%	0.1		
95%	0.05	0.025	1.96
98%	0.02	0.01	2.33
99%			2.575
99.74%			3
86.64%	0.1336	0.0668	1.5

Attendance Survey Question

- On a 4”x6” index card
  - Please write down your name and section number
  - Today’s Question: Are you going to watch the NCAA Basketball game this weekend?

a. All

b. Some

c. None

Facts About Confidence Intervals I

- The width of a confidence interval
  - Increases as the confidence level increases
  - Increases as the error probability decreases
  - Increases as the standard error increases
  - Decreases as the sample size *n* increases