Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. When to use t* vs. 1.96 for CIs.
2. Bradley Effect example.
3. More about confounding factors.
4. Lefties example.

Read chapter 5.

HW2 is due Wed, Feb12, 1159pm. 2.3.15, 3.3.18, and 4.1.23.

Midterm is Mon Feb24 in class.

The course website is http://www.stat.ucla.edu/~frederic/13/W25 .

1. Formulas for CIs for one variable, quantitative or categorical.

if the observations are iid and n is large, then

P(μ is in the range $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$) ~ 95%.

and since s ~ σ when n is large, 95% CI is

$\bar{x}$ +/- 1.96 $s/\sqrt{n}$.

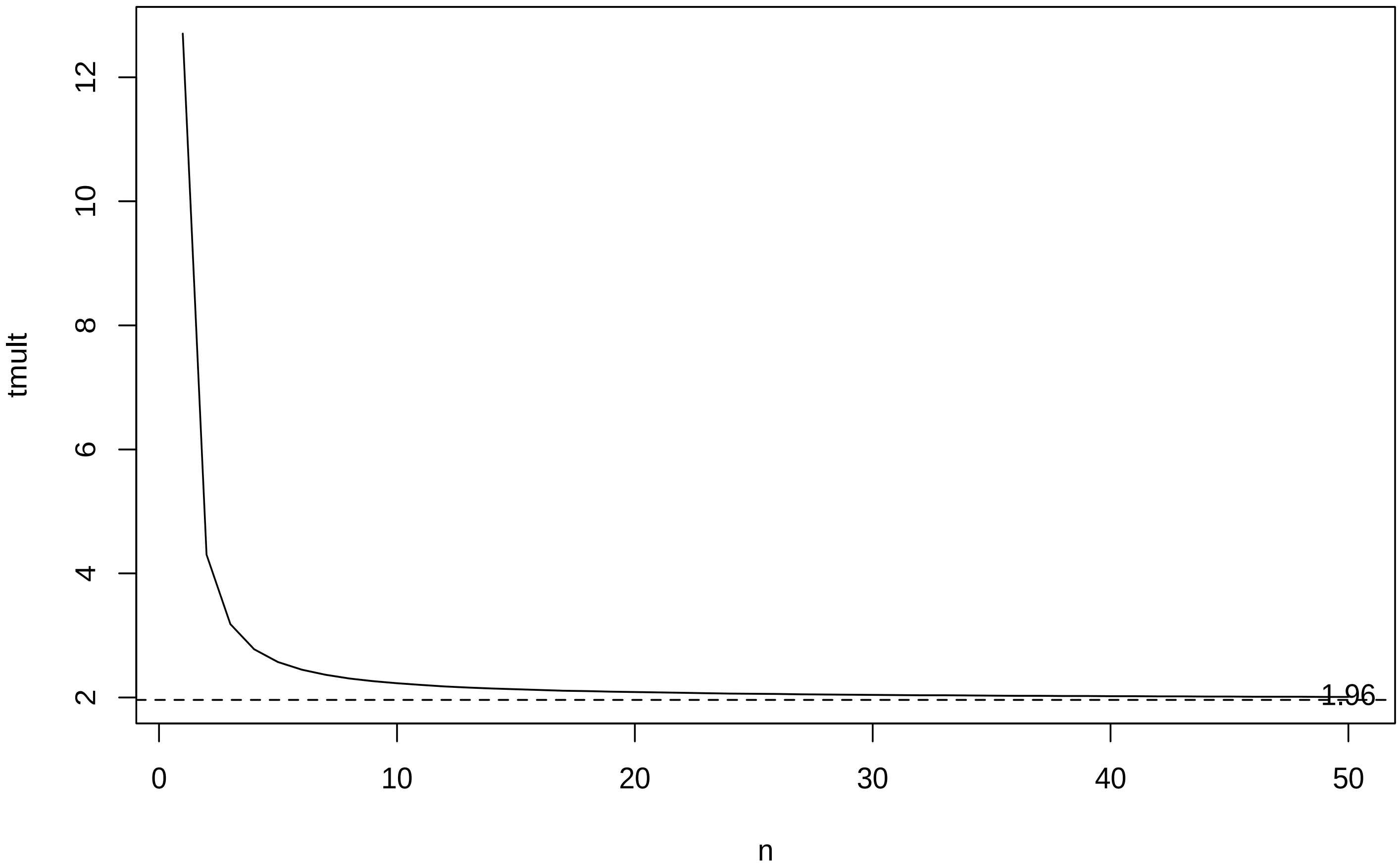
If the obs. are iid and normal and σ is unknown, then

even if n is small,

P(μ is in the range $\bar{x}$ +/- t_{mult} $s/\sqrt{n}$) ~ 95%.

where t_{mult} depends on n.

$\bar{x}$ +/- t_{mult} $s/\sqrt{n}$.



t_{mult} gets really close to 1.96 when n gets larger than about 30, so for this class we will use the rule of thumb

$n \geq 30$ is large, for quantitative data. For categorical, at least 10 of each type in your sample will be the rule of thumb.

Formulas for CIs for one variable, quantitative or categorical.

Note that for quantitative variables, in the 95% CI formula

$$\bar{x} \text{ +/- } 1.96\ s/\sqrt{n}\ ,$$

The quantity $s / \sqrt{n}$ is called the SE for the mean.

For categorical data, the population is never normal!
View the values as 0 or 1. Then

$\hat{p} = \bar{x}$, and $s = \sqrt{[\hat{p}(1-\hat{p})]}$. So the formula for a 95% CI is

$$\hat{p} \text{ +/- } 1.96 \sqrt{[\hat{p}(1-\hat{p})/n]}.$$

Here large n means ≥ 10 of each type in the sample.

2. Cautions When Conducting Inference, and the controversial "Bradley Effect"

Example 3.5A

The “Bradley Effect”

- Tom Bradley, long-time mayor of Los Angeles, ran as the Democratic Party’s candidate for Governor of California in 1982.
  - Political polls of likely voters showed Bradley with a significant lead in the days before the election.
  - Exit polls favored Bradley significantly.
  - Many media outlets projected Bradley as the winner.
- Bradley narrowly lost the overall race.

The “Bradley Effect”

- After the election, research suggested a smaller percentage of white voters had voted for Bradley than polls predicted.
- A very large proportion of undecided voters voted for Deukmejian.

The "Bradley Effect"

- What are explanations for this discrepancy?
  - Likely voters answered the questions with a "social desirability bias".
  - They answered polling questions the way they thought the interviewer wanted them to.
- Discrepancies in polling and elections has since been called the "Bradley effect".
- It has been cited in numerous races and has included gender and other stances on political issues.

Clinton vs. Obama

- In the 2008 New Hampshire democratic primary
  - Obama received 36.45% of the primary votes.
  - Clinton received 39.09%.
- This result shocked many since Obama seemed to hold a lead over Clinton.
- USA Today/Gallup poll days before the primary, n = 778.
  - 41% of likely voters said they would vote for Obama.
  - 28% of likely voters said they would vote for Clinton.
- How unlikely are the Clinton and Obama poll numbers given that 39.09% and 36.45% of actual primary voters voted for Clinton and Obama?

Clinton vs. Obama

- We’re assuming that the 778 people in the survey are a good representation of those who will vote.
  - The 778 people aren’t a simple random sample.
- Pollsters used random digit dialing and asked if respondents planned to vote in the Democratic primary.
  - 9% (a total of 778) agreed to participate.
  - 319 said that they planned to vote for Obama and 218 for Clinton.

Clinton vs. Obama

Suppose we make the following assumptions:

1. Random digit dialing is a reasonable way to get a sample of likely voters.
2. The 9% who participated are like the 91% who didn’t.
3. Voters who said they planned to vote actually voted in the primary.
4. Answers to who they say they will vote for match who they actually vote for.

Then we expect the sample proportion roughly to agree with the final vote proportion.

Clinton vs. Obama

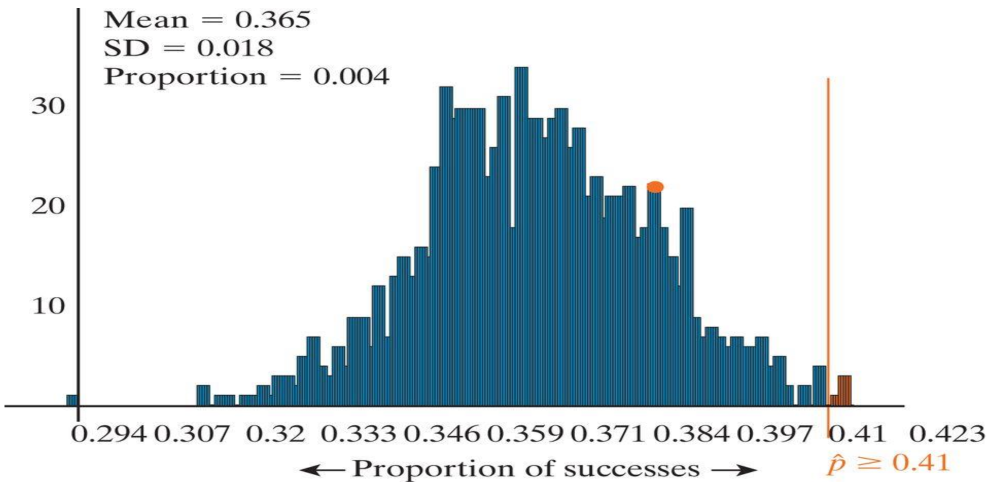
- One question is whether the proportion of likely voters who say they will vote for Obama is the same as the proportion of likely voters who actually vote for Obama (observed on primary day to be 0.3645).
- What would the Bradley Effect do in this case?
  - The proportion who say they will vote for Obama would be larger than 0.3645.

Clinton vs. Obama

- State the Null and Alternative hypotheses.
  - Null: The proportion of likely voters who would claim to vote for Obama is 0.3645.
  - Alternative: The proportion of likely voters who would claim to vote for Obama is higher than 0.3645.

Clinton vs. Obama

- Simulation of 778 individuals randomly chosen from a population where 36.45% vote for Obama
- The chance of getting a sample proportion of 0.41 successes or higher is very small. 0.004.



Clinton vs. Obama

- Convincing evidence that the discrepancy between what people said and how they voted is not explained by random chance alone.
- At least one of the 4 model assumptions is not true.

Clinton vs. Obama

1. **Random digit dialing is a reasonable way to get a sample of likely voters**
   – Roughly equivalent to a SRS of New Hampshire residents who have a landline or cell phone
   – Slight over-representation of people with more than one phone

Clinton vs. Obama

2. **The 9% of individuals reached by phone who agree to participate are like the 91% who didn't**
   – 91% includes people who didn't answer their phone and who didn't participate
   – Assumes that respondents are like non-respondents.
   – The ***response rate*** was very low, but typical for phone polls
   – No guarantee that the 9% are representative.

Clinton vs. Obama

3. **Voters who said they plan to vote in the Democratic primary will vote in the primary.**
   – There is no guarantee.
4. **Respondent answers match who they actually vote for.**
   There is no guarantee.

Clinton vs. Obama

Because of the wide disparity between polls and the primary, an independent investigation was done with the following conclusions:

1. People changed their opinion at the last minute
2. People in favor of Clinton were more likely not to respond
3. The Bradley Effect
4. Clinton was listed before Obama on every ballot

These are examples of **nonrandom errors.**

3. More about confounding factors.

- By a confounding factor, we mean an alternative explanation that could explain the apparent relationship between the two variables, even if they are not causally related. Typically this is done by finding another difference between the treatment and control group. For instance, different studies have examined smokers and non-smokers and have found that smokers have higher rates of liver cancer. One explanation would be that smoking causes liver cancer. But is there any other, alternative explanation?
- One alternative would be that the smokers tend to drink more alcohol, and it is the alcohol, not the smoking, that causes liver cancer.

More about confounding factors.

- Another plausible explanation is that the smokers are probably older on average than the non-smokers, and older people are more at risk for all sorts of cancer than younger people.
- Another might be that smokers engage in other unhealthy activities more than non-smokers.
- Note that if one said that “smoking makes you want to drink alcohol which causes liver cancer,” that would not be a valid confounding factor, since in that explanation, smoking effective is causally related to liver cancer risk.

4. Lefties example.

- A confounding factor must be plausibly linked to both the explanatory and response variables. So for instance saying "perhaps a higher proportion of the smokers are men" would not be a very convincing confounding factor, unless you have some reason to think gender is strongly linked to liver cancer.
- Another example: left-handedness and age at death. Psychologists Diane Halpern and Stanley Coren looked at 1,000 death records of those who died in Southern California in the late 1980s and early 1990s and contacted relatives to see if the deceased were righthanded or lefthanded. They found that the average ages at death of the lefthanded was 66, and for the righthanded it was 75. Their results were published in prestigious scientific journals, Nature and the New England Journal of Medicine.

Lefties example.

All sorts of causal conclusions were made about how this shows that the stress of being lefthanded in our righthanded world leads to premature death.

The New York Times

U.S.

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

POLITICS EDUCATION TEXAS

Being Left-Handed May Be Dangerous To Life, Study Says

Reuters
Published: April 4, 1991

BOSTON, April 3— Left-handed people tend to live significantly shorter lives than right-handers, perhaps because they face more perils in a world dominated by the right-handed, according to new research.

FACEBOOK
TWITTER
GOOGLE+

Lefties example.

- Is this an observational study or an experiment?

Lefties example.

- Is this an observational study or an experiment?

It is an observational study.

- Are there plausible confounding factors you can think of?

Lefties example.