Warm-up

What conditions must we check before we can use the ~~Normal curve~~ to approximate the sampling distribution for proportions? → % → p ? $\hat{p}$

Randomization
10% Condition
Success/Failure $n \cdot p \geq 10$ $n \cdot q \geq 10$

SD $\sqrt{\frac{pq}{n}}$

mean/average:

Randomization
10%
Large enough

$SD = \frac{\sigma}{\sqrt{n}}$

Jan 17-9:33 PM

On a piece of paper, index card, etc. Answer the following:

(a) Best known for its testing program, ACT Inc., also compiles data on a variety of issues in education. In 2004 the company reported that the national college freshman-to-sophomore retention rate held steady at 74% over the previous four years. Consider random samples of 400 freshmen who took the ACT. Use the 68-95-99.7 Rule to describe the sampling distribution model for the percentage of those students we expect to return to that school for their sophomore years. Do you think the appropriate conditions are met?

(b) Based on the 74% national retention rate described above, does a college where 522 out of 603 freshman returned the next year as sophomores have a right to brag that it has an unusually high retention rate? Explain.

Jan 28-9:13 AM

Chapter 19
Confidence Intervals for Proportions

Read Chapter 19!

%

Jan 17-9:35 PM

Standard Error

Both of the sampling distributions we've looked at are Normal.

- For proportions

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}}$$

$\hat{p}$ = sample proportion

p = pop. proportion

- For means

$$SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$$

pop. st.dev

$\bar{y}$ or $\bar{x}$ = sample mean

μ = pop. mean

Jan 17-9:35 PM

Standard Error Cont'd

pop. st. dev. (mean)

- When we don't know p or σ, we're stuck, right?

population proportion

samples $\hat{p}$, $\bar{x}$

- Nope. We will use sample statistics to estimate these population parameters.

population μ, σ

- Whenever we estimate the standard deviation of a sampling distribution, we call it a standard error.

Jan 17-9:37 PM

Standard Error Cont'd

- For a sample proportion, the standard error is

sample prop.

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$\hat{p}$ = sample prop

$\hat{q} = 1 - \hat{p}$

- For the sample mean, the standard error is

$$SE(\bar{y}) = \frac{sd(\bar{y})}{\sqrt{n}}$$

sample st. dev.

$\sigma \rightarrow$ pop. st. dev.

$s \leftarrow$ sample

Jan 17-9:38 PM

A Confidence Interval

- Recall that the sampling distribution model of $\hat{p}$ is centered at p, with standard deviation

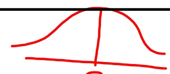
$\sigma = \sqrt{\frac{pq}{n}}$ $N\left(p, \sqrt{\frac{pq}{n}}\right)$

- Since we don't know p, we can't find the true standard deviation of the sampling distribution model, so we need to find the standard error:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Jan 17-9:39 PM

A Confidence Interval



By the 68-95-99.7% Rule, we know

P

- about 68% of all samples will have $\hat{p}$'s within 1 SE of p

sample proportions

- about 95% of all samples will have $\hat{p}$'s within 2 SEs of p

- about 99.7% of all samples will have $\hat{p}$'s within 3 SEs of p

We can look at this from $\hat{p}$'s point of view.

Jan 17-9:42 PM

A Confidence Interval Cont'd

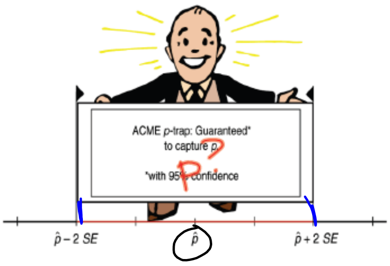
Consider the 95% level:

- There's a 95% chance thatp is no more than 2 SEs away from $\hat{p}$
- So, if we reach out 2 SEs, we are 95% sure that p will be in that interval. In other words, if we reach out 2 SEs in either direction of $\hat{p}$, we can be 95% confident that this interval contains the true proportion.

This is called a 95% confidence interval.

Jan 17-9:45 PM

A Confidence Interval Cont'd



Reaching out 2 *SE*s on either side of $\hat{p}$ makes us 95% confident we'll trap the true proportion, *p*. **Figure 19.2**

Jan 17-9:48 PM

TI Tips: Finding Confidence Intervals

- Let's say we have a sample of 104 cows and 54 were diseased.
- STAT
- TESTS
- (Our proportion is based on one sample with a Normal model) A: 1-PropZInt
- Enter the number of successes observed (x). Must be an integer.
- Enter the sample size (n).
- Specify a confidence level (usually .95)
- Calculate
- "Based on my sample, I am 95% confident that between 42.3% and 61.5% of all cows will be diseased."

Jan 17-9:48 PM

A pew research study regarding cell phones asked questions about cell phone experience. One growing concern is unsolicited advertising in the form of text messages. Pew asked cell phone owners, "Have you ever received unsolicited text messages on your cell phone from advertisers?" and 17% reported that they had. Pew estimates a 95% confidence interval to be 0.17 plus or minus 0.04, or between 13% and 21%.
Are the following correct? Explain.

$\hat{p} \pm \underbrace{ME}_{2SE}$

1. ~~In Pew's sample~~, somewhere between 13% and 21% of respondents reported that they had received unsolicited advertising text messages. False
2. We can be 95% confident that 17% ($\hat{p}$) of US cell phone owners have received unsolicited advertising text messages. False
3. We are 95% confident that between 13% and 21% of all US cell phone owners have received unsolicited advertising text messages. True
4. We know that between 13% and 21% of all US cell phone owners have received unsolicited advertising text messages. False
5. 95% of all US cell phone owners have received unsolicited advertising text messages. False

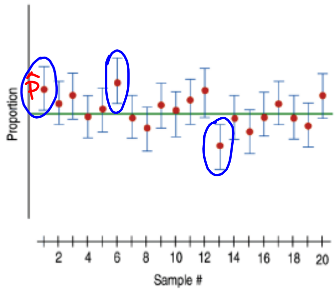
CI

Jan 1-11:17 PM

What Does "95% Confidence" Really Mean?

- Each confidence interval uses a sample statistic ($\bar{x}$ or $\hat{p}$) to estimate a population parameter. p
- But, since samples vary, the statistics we use, and thus the confidence intervals we construct, vary as well.
  - The figure to the right shows that some of our confidence intervals (from 20 random samples) capture the true proportion (the green horizontal line), while others do not.

p



Jan 17-10:04 PM

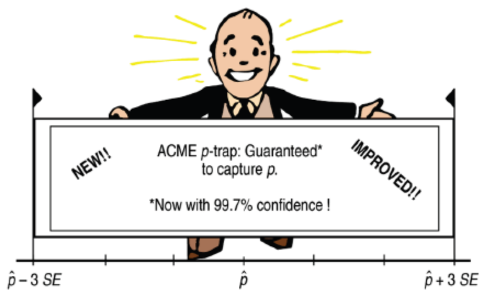
What Does "95% Confidence" Really Mean?

- Our confidence is in the process of constructing the interval, not in any one interval itself.
- Thus, we expect 95% of all 95% confidence intervals to contain the true parameter that they are estimating.

Jan 17-10:08 PM

Margin of Error: Certainty vs. Precision

- We can claim, with 95% confidence, that the interval $\hat{p} \pm \underbrace{2SE(\hat{p})}_{\text{Margin of error}}$ contains the true population proportion. Margin of error (will change w/ conf. level)
  - The extent of the interval on either side of $\hat{p}$ is called the margin of error (*ME*).
- In general, confidence intervals have the form *estimate ± ME*.
- The more confident we want to be, the larger our *ME* needs to be, making the interval wider.



Jan 17-10:11 PM

Margin of Error: Certainty vs. Precision

- To be more confident, we wind up being less precise.
  - We need more values in our confidence interval to be more certain.
- Because of this, every confidence interval is a balance between certainty and precision.
- The most commonly chosen confidence levels are 90%, 95%, and 99% (but any percentage can be used).

Jan 17-10:12 PM

Example: On January 30-31, 2007, Fox News polled 900 registered voters nationwide. When asked, "Do you believe global warming exists?" 82% said "yes." Fox reported their margin of error to be +/-3%. It is standard among pollsters to give a 95% confidence level unless otherwise stated. Given that, what does Fox news mean by claiming a margin of error of +/-3% in this context?

The true pop. proportion of belief in Global warming will be within (±) 3% of 95% of our 95% CI.

Example: It is a convention among pollsters to use a 95% confidence level and to report the "worst case" margin of error, based on p = .5. How did Fox calculate their margin of error?

$\boxed{ME = 2SE}$

$= 2\left(\sqrt{\frac{pq}{n}}\right) = 2\left(\sqrt{\frac{(.5)(.5)}{900}}\right)$

$= 2(.017)$

$= .033$

Jan 17-10:09 PM

Critical Values

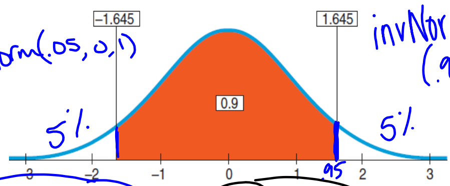
- The '2' in $\hat{p} \pm 2SE(\hat{p})$ (our 95% confidence interval) came from the 68-95-99.7% Rule.
- Using a table or technology, we find that a more exact value for our 95% confidence interval is 1.96 instead of 2.
  - We call 1.96 the critical value and denote it z^*.
- For any confidence level, we can find the corresponding critical value (the number of SEs that corresponds to our confidence interval level).

95% → w/i 2 SD
68% → w/i 1 SE

Example: For a 90% confidence interval, the critical value is 1.645:

invNorm(.05, 0, 1)

invNorm(.95, 0, 1)



5% 5% .95

$z^*_{.90} = 1.645$

$z^*_{.95} = 1.96$
invNorm(.9750, 1)

Jan 17-10:15 PM

$z^*_{.75} = 1.15$

.125 .75 12.5

invNorm(.125)
invNorm(.875, 0, 1)
1 − .125

$z^*_{.99} = 2.576$

5% 99% .005 .995

Dec 12-12:16 PM

85% CI

7.5 85% 7.5 85%

invNorm(.925, 0, 1)
$z^* = 1.44$

100
− 7.5
92.5

invNorm(.075)
− 1.44

52% Confidence

24% 52% 24%

100
− 24

invNorm(.76, 0, 1) = 0.706
(.24, 0, 1) z^*

$z^*_{.52} = .706$

$z^*_{.85} = 1.44$

invNorm(.995, 0, 1)

$z^*_{.95} = 1.96$ $z^*_{.9} = 1.645$

$z^*_{.99} = 2.576$

Jan 4-1:39 PM

Example: The Fox news poll of 900 registered voters found that 82% of the respondents believed that global warming exists. Fox reported a 95% confidence interval with a margin of +/-3%. using the critical value of z and the standard error based on the observed proportion, what would be the margin of error for a 90% confidence interval? What's good and bad about this change?

$ME_{.90} = \left(\underset{z^*}{\text{Critical value}}\right) SE$

$= (1.645)\left(\sqrt{\frac{(.82)(.18)}{900}}\right) = \pm 2.1\%$

more precise but less confident
(smaller interval)

Jan 17-10:20 PM

Example 1: If Fox wanted to be 98% confident, would their confidence interval need to be wider or narrower?

Wider

Example 2: Fox's margin of error was about +/-3%. If they reduced it to +/-2%, would their level of confidence be higher or lower?

lower

Example 3: If Fox News had polled more people, would the interval's margin of error have been larger or smaller?

$95\% \rightarrow n = 2000$

$\underset{z^*}{(1.96)}\left(\underset{SE}{\sqrt{\frac{(.82)(.18)}{2000}}}\right) = 1.68\%$ ME

Jan 17-10:21 PM

Assumptions and Conditions

- The assumptions and the corresponding conditions you must check before creating a confidence interval for a proportion are:
- Independence Assumption: we first need to think about whether the Independence Assumption is plausible. It's not one you can check by looking at the data. Instead, we check two conditions to decide whether independence is reasonable. Independence
  - Randomization Condition: Were the data sampled at random or generated from a properly randomized experiment? Proper randomization can help ensure independence.
- Sample Size Assumption: The sample needs to be large enough for us to be able to use the CLT.
  - -10% Condition: Is the sample size no more than 10% of the population?
  - -Success/Failure Condition: We must expect at least 10 "successes" and at least 10 "failures." $n\hat{p} \geq 10$ & $n\hat{q} \geq 10$

Jan 17-10:23 PM

Warm-up

What is the formula for Margin of error? $2\,SE \rightarrow 95\%$

(Critical value × SE)

How do we set up a confidence interval?

Sample statistic ± (Critical value)(SE)

$\hat{p} \pm z^* SE(\hat{p})$

What is the Z* for an 92% confidence level?

1.75

invNorm(.96)

Jan 5-10:38 AM

ONE-PROPORTION z-INTERVAL

When the conditions are met, we are ready to find the confidence interval for the population proportion, p. The confidence interval is $\hat{p} \pm z^* \times SE(\hat{p})$ where the standard deviation of the proportion is estimated by $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$.

- The critical value, z^*, depends on the particular confidence level, C, that you specify.

Jan 17-10:26 PM

Example: Direct mail advertisers send solicitations (aka "junk mail") to thousands of potential customers in the hope that some will buy the company's product. The acceptance rate is usually quite low. Suppose a company wants to test the response to a new flyer, and sends it to 1000 people randomly selected from their mailing list of over 200,000 people. They get orders from 123 of the recipients.

(a) Create a 90% confidence interval for the percentage of people the company contacts who may buy something.

(b) Explain what this interval means.

(c) Explain what "90% confidence" means.

(d) The company must decide whether to now do a mass mailing. The mailing won't be cost effective unless it produces at least a 5% return. What does your confidence interval suggest? Explain.

Conditions:

randomization: Flyer is sent to 1000 randomly selected people.

10%: 1000 is less than 10% of the 200,000 people on their mailing list.

S/F: $n \cdot \hat{p} \geq 10$ $(1000)(.123)$ $123 \geq 10$ ✓ $\quad n \cdot \hat{q} \geq 10$ $1000(.877)$ $877 \geq 10$ ✓

$\hat{p} = \frac{123}{1000}$

*Since conditions are met, we can continue with a one-proportion confidence interval using normal model.

$\hat{p} = .123$

$\hat{q} = .877$

$SE = \sqrt{\frac{(.123)(.877)}{1000}}$

(90%)

$z^*_{.9} = 1.645$ $\quad SE(\hat{p}) = 0.0104$

$\hat{p} \pm z^* SE(\hat{p})$

$(.123) \pm (1.645)(0.0104)$ $\quad (.106, .14)$

We are 90% confident that the proportion of people who will buy something from the company falls between 10.6% and 14%.

Jan 4-9:28 AM

(b) We are 90% confident, based on this sample, that the proportion of people contacted who may buy something is between 10.6% and 14.0%.

(c) About 90% of all random samples will produce confidence intervals that contain the true population proportion.
90% of the 90% CIs will contain the true pop. prop.

(d) Do the mass mailing. The interval is considerably above 5%.

Jan 4-9:41 AM

Choosing Your Sample Size

- To determine how large a sample to take:
- Choose a Margin of Error (ME) and a Confidence Interval Level.
- The formula requires $\hat{p}$ which we don't have yet because we have not taken the sample. A good estimate for $\hat{p}$ which will yield the largest value for pq (and therefore n) is 0.50.
- Solve the formula for n. (to be safe, round up n)

$$\boxed{ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

$$z^* \sqrt{\frac{(.5)(.5)}{n}}$$

(Critical Value)(SE)

Jan 17-10:29 PM

Example: The Fox poll estimated that 82% of all voters believed in global warming exists with a margin of error of +/-3%. Suppose an environmental group planning a follow-up survey of voters' opinions on global warming wants to determine a 95% confidence interval with a margin of error no more than +/-2%. How large a sample do they need? Use the Fox news estimate as the basis for your calculation.

$\hat{p} = .82$

$\hat{q} = .18$

$$ME = z^* SE(\hat{p})$$

$$\frac{.02}{1.96} = \frac{(1.96)}{1.96}\sqrt{\frac{(.82)(.18)}{n}}$$

$$.0102^2 = \sqrt{\frac{(.82)(.18)}{n}}^{\,2}$$

$$n \cdot .000104 = \frac{(.82)(.18)}{\cancel{n}} \cdot \cancel{n}$$

$$\frac{\cancel{(.0001)}n}{\cancel{.0001}} = \frac{(.82)(.18)}{.000104}$$

$$n = \frac{(.82)(.18)}{.000104} = 1419 \text{ people}$$

Jan 17-10:31 PM

Warm-up

A credit card company is about the send out a mailing to test the market for a new credit card. From that sample, they want to estimate the true proportion of people who will sign up the card nationwide. A pilot study suggests that about 0.5% of the people receiving the offer will accept it. To be within a tenth of a percentage point (0.001) of the true rate with 95% confidence, how big does the test mailing have to be?

Jan 17-10:33 PM

Today's Practice pg. 455

Ch. 19, #1, 3 (parts a & b only), 5, 7, 9, 11, 13, 15, 22, 27, 29, 31

Jan 17-11:05 PM

Warm-up

You are going to create a 95% confidence interval for a population proportion and want the margin of error to be no more than 0.05. Historical data indicate that the population proportion has remained constant at about 0.7. What is the minimum size random sample you need to construct this interval?

A. 385
B. 322
C. 274
D. 275
E. 323

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

95% confidence

$ME = .05$

$\hat{p} = .7$

$\hat{q} = .3$

n ?

$$.05 = 1.96\sqrt{\frac{(.7)(.3)}{n}}$$

$$\frac{.05}{1.96} = \sqrt{\frac{.21}{n}}$$

$$n \cdot \left(\frac{.05}{1.96}\right)^2 = \frac{.21}{\cancel{n}} \cdot \cancel{n}$$

$$n = \frac{.21}{\left(\frac{.05}{1.96}\right)^2}$$

322.6

323

*Read Ch. 20!

Jan 17-11:05 PM

Dec 12-7:58 AM