Math 140
Introductory Statistics

Next midterm May 1

8.1 Confidence intervals

54% of Americans approve the job the president is doing with a margin error of 3%

55% of 18-29 year olds consider themselves unattached according to a poll with margin error of 3%

51% of Americans assign a grade of A or B to the public schools in their community.
This survey had a margin of error of 3%

What does this mean?

51% of Americans assign a grade of A or B to the public schools in their community.

This survey had a margin of error of 3%

These results are based on telephone interviews with a randomly selected sample of 1108 adults, conducted May 23–June 6, 2001

This means that we are 95% confident that if we were to ask ALL AMERICANS about their schools, we would find that the real value would be somewhere between 51-3% and 51+3%, that is 48% and 54%

What does this mean?

We can also say that the 48%-54% values
Are plausible values for the proportion p
of people who approve of US schools.

Single people in the US

Of the 1068 young singles surveyed, 55% are unattached.

If we asked ALL young people in the US we would know
The EXACT value of p.

For the sample surveyed all we have is $\hat{p}$ = 55%

Determine whether this value for $\hat{p}$ lies in the
middle 95% of the sampling
distribution for the proposed values of p = 0.57 and 0.51

If p = 0.57

Then, we know that the standard deviation for the entire population would be

$$\sigma = \sqrt{p(1-p)} = \sqrt{0.57 * 0.43} = 0.495$$

Then, the standard error for our sample of n=1086 people would be

$$\sigma_p = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = \frac{0.495}{\sqrt{1068}} = 0.015$$

If p = 0.57

And the 95% confidence interval would be between

0.57 - 1.96 * 0.015 = 0.54 and 0.57 + 1.96 * 0.015 = 0.60

$$p - 1.96 * \sigma_p \qquad\qquad p + 1.96 * \sigma_p$$

This means that if p = 0.57, then the 95% confidence interval is between 0.54 and 0.60

If p = 0.57

We measured from our sample $\hat{p} = 55\%$

So we can conclude that since the value we measured (0.55) falls between the values 0.54 and 0.60,

57% is a plausible value for the true p.

Basically, you have to do it "backwards"

Take the proposed true value of p

Calculate the 95% confidence interval associated to your sample size

Check whether your estimate from the sample you have of $\hat{p}$ is compatible with the 95% interval you have just measured.

Basically, you have to do it "backwards"

And then if your measured value of $\hat{p}$
falls within that interval
The proposed p is plausible.

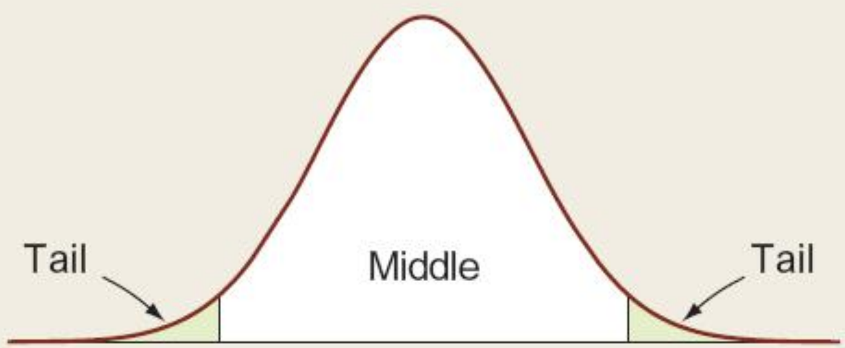
If not, p is not plausible

Remember, p is a guess, $\hat{p}$ is "partial" data.

What does this mean?

Plausible Values for p

You have a random sample from a population with an unknown proportion of successes, p. If the sample proportion, $\hat{p}$, falls in the middle of the sampling distribution of $\hat{p}$ for the population with proportion of successes p_0, then p_0 is a plausible value for the proportion of successes in the population. If $\hat{p}$ falls in the tails, then p_0 is not plausible as the proportion of successes in the population. "Middle" typically refers to the middle 95%, but may be 90%, 99%, or other values depending on the situation.



Try the same with p=0.51

Is p=0.51 plausible?

Recall, you measured on n=1068 people
And found that $\hat{p} = 0.55$

Calculate the proposed standard deviation,
The proposed standard error
And then the proposed 95% confidence interval

Try the same with p=0.51

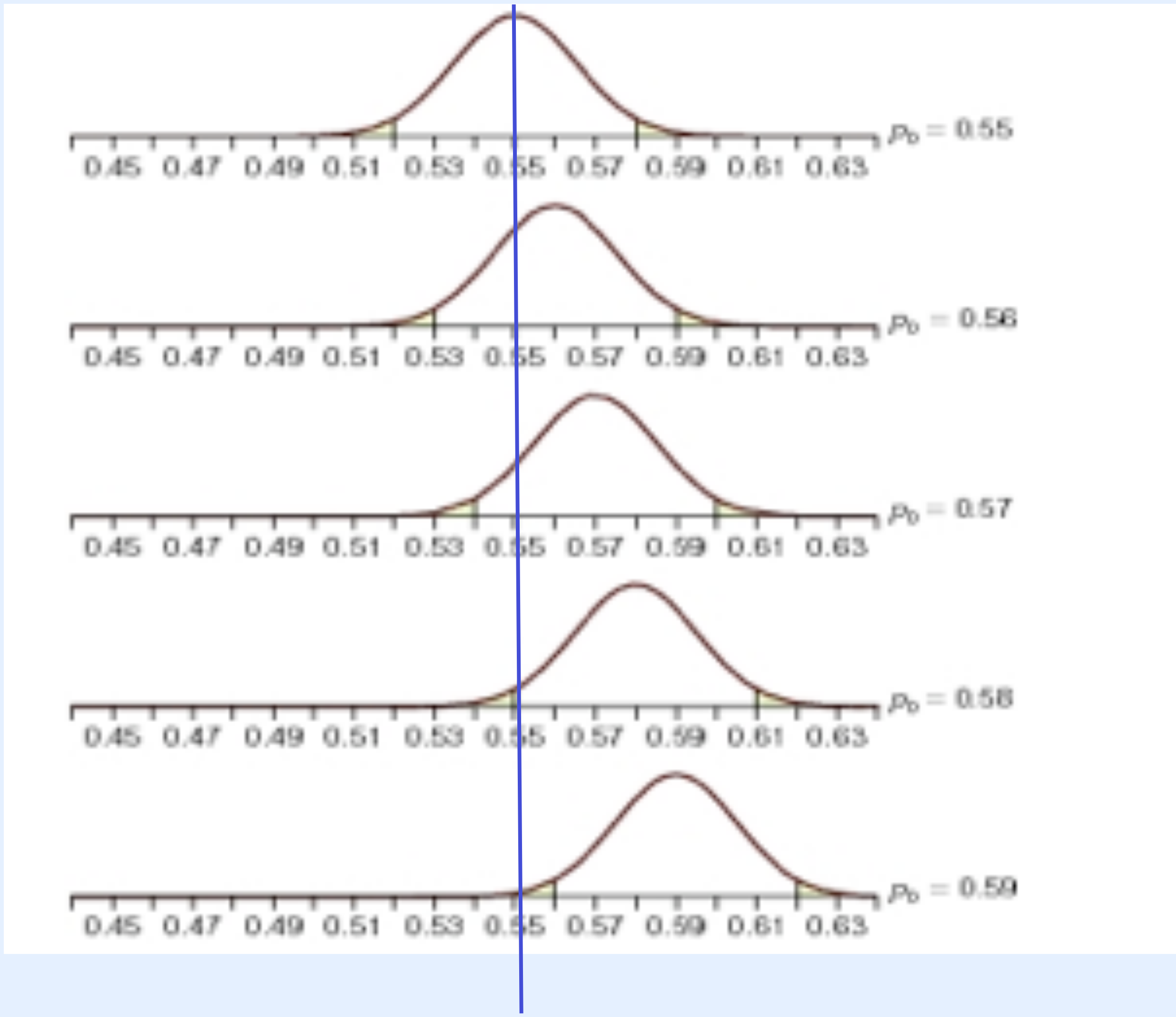
You should find that the value of

$$\hat{p} = 0.55$$

falls outside the 95% interval confidence of the proposed true value of p = 0.51

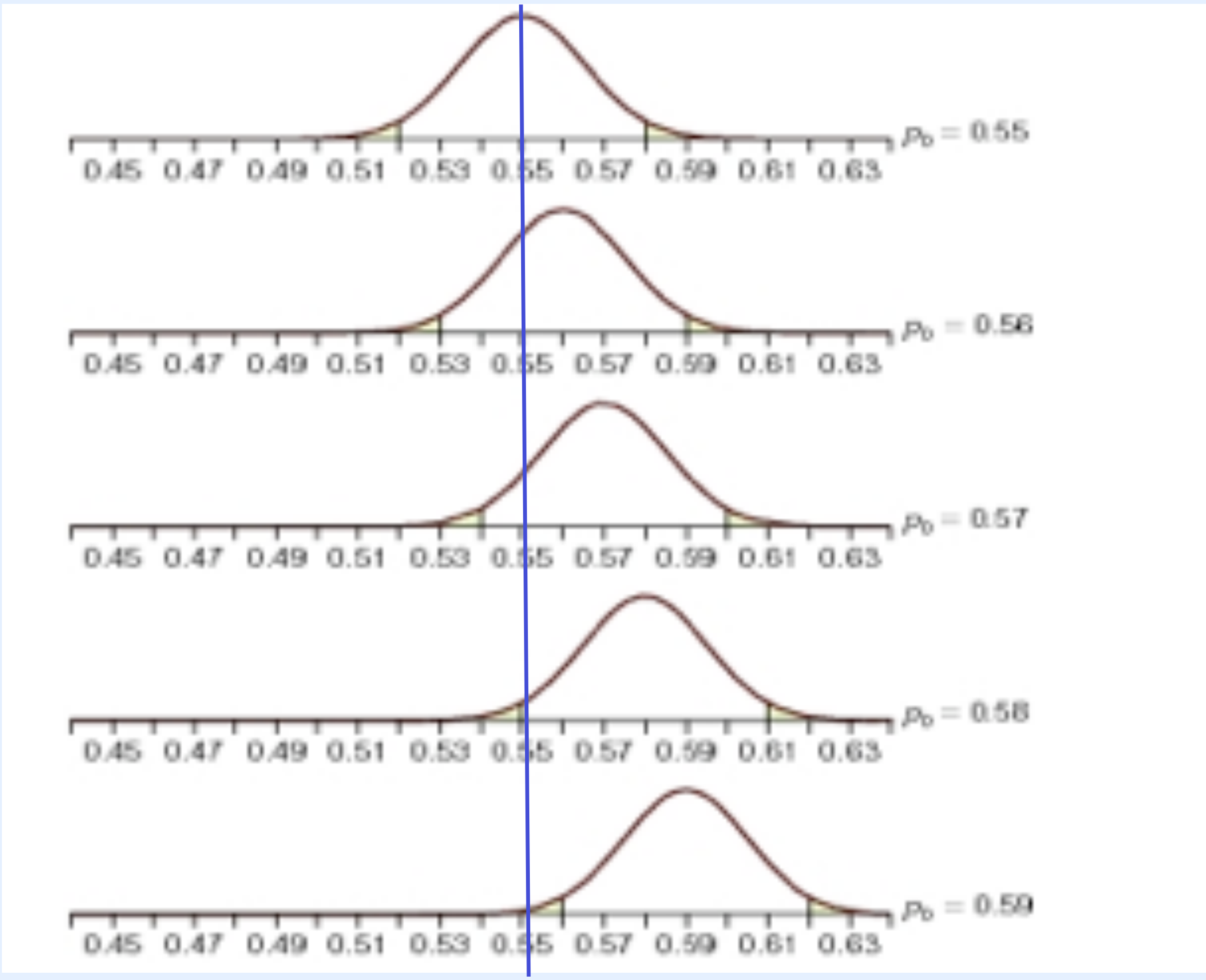
This means that p=0.51 is not a plausible value
For the true value of p.

Guesses from p=0.55 to p=0.59



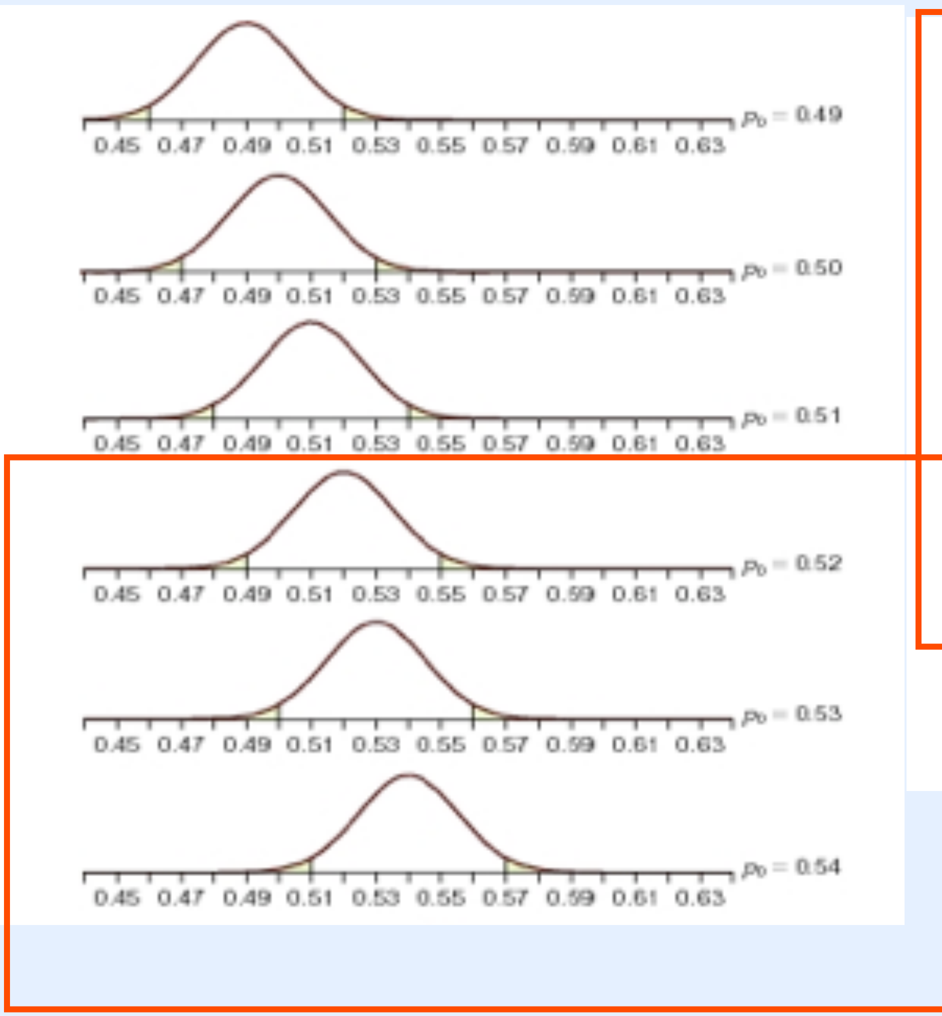
This is my measured value of 0.55

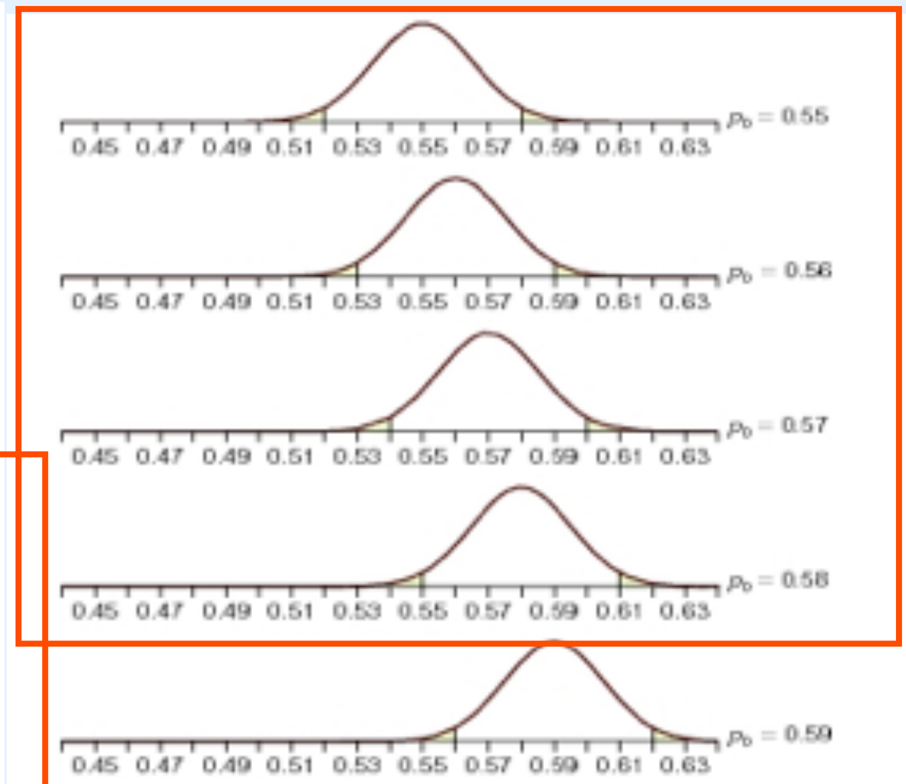
Guesses from p=0.55 to p=0.59



They are all compatible with the 95% confidence interval except the last one

Guesses from p=0.49 to p=0.59

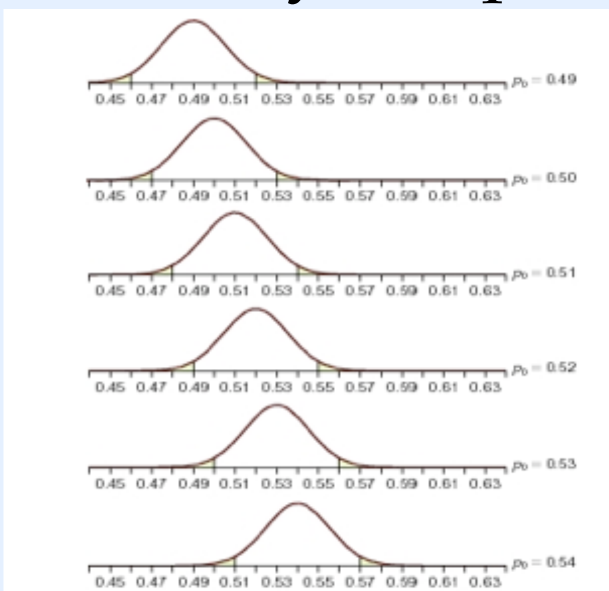


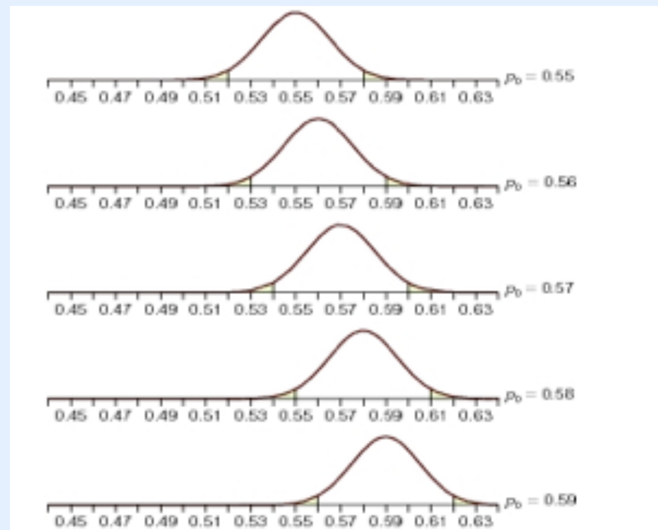


In the red boxes the plausible values

The confidence interval

Contains all the values of p that are plausible given my sample measurement of $\hat{p}$





Here it would be between 0.52 and 0.58

The confidence interval

Contains all the values of p that are plausible given my sample measurement of $\hat{p}$

A good estimate for all my plausible values of p is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where z* = 1.96 for the 95% confidence level
or z* = 1.646 for the 90% confidence level

Rule of thumb

Can use if n$\hat{p}$ and n(1-$\hat{p}$) are at least 10

The size of the population is at least
10 times that of the sample

The sample is random, success probability statistics

$$\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

For you

313 students were surveyed about their sleep patterns. It was found that 43% woke up at least once a night.

What is the population? What is the parameter we are studying?

Build the 95% confidence interval for all college students that woke up at least once a night.

For you

313 students were surveyed about their sleep patterns. It was found that 43% woke up at least once a night.

Our population are the Students at our university - proportion of students who wake up at least once

Build the 95% confidence interval for all college students that woke up at least once a night.

For you

313 students were surveyed about their sleep patterns. It was found that 43% woke up at least once a night.

$$\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.43 \pm 1.96 * \sqrt{\frac{0.43 * 0.57}{313}} = 0.43 \pm 0.055$$

$$0.43 + 0.055 = 0.485$$

$$0.43 - 0.055 = 0.375$$

check that np and n(1-p) are at least 10

So, the 95% confidence interval is for all p values between 0.375 and 0.485

Our conclusion is that

If we had asked EVERYONE on campus

then with 95% confidence we would
have been sure that
between 37.5% and 48.5% of them woke up at least
once a night.

Margin of error

$$z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Half of the confidence interval

Back to being single

55% people reported they were single.
The margin of error that we can estimated is

$$z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 * \sqrt{\frac{0.55 * 0.45}{1068}} = 0.0298$$

So we can say that the margin of error is about 2.98 or 3%

College life in the US

184,457 surveys from first year students in the US
722 colleges

17% report spending more than 20 hours a week "studying"

Build the 95% confidence interval for the proportion of students who studied at least 20 hours a week.

College life in the US

184,457 surveys from first year students in the US
722 colleges

17% report spending more than 20 hours a week "studying"

Build the 95% confidence interval for the proportion of students who studied at least 20 hours a week.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

College life in the US

184,457 surveys from first year students in the US
722 colleges

17% report spending more than 20 hours a week "studying"

We should be 95% confident that the proportion of Students who studied more than 20 hours a week is between
0.1683 and 0.1717

College life in the US

194,858 surveys from senior students in the US
722 colleges

20% report spending more than 20 hours a week "studying"

Calculate the same 95% confidence interval

College life in the US

Freshmen: 95% confidence interval
0.1683 and 0.1717

Seniors: 95% confidence interval
0.1982 and 0.2018

There is no overlap and we can say
That seniors do study more than freshmen!

The capture rate

We have 200 samples of students who are asked if they borrow money to attend college

From each of the samples we construct the 95% confidence interval
(finding the mean and SD of their data)

If the true value of people who borrow money is 53%
Then we expect that of those 200 samples,
95% of them will contain the value 0.53

The capture rate is 95% = 95*200 = 190 samples

The capture rate

55% of young people are single
Random sample of 1068 people
Margin error 3%
Confidence Interval from 52% to 58%

If we would ask all young people if they were single, we are 95% certain that between 52% and 58% would say yes.

The capture rate

55% of young people are single

Random sample of 1068 people

What is the 99% confidence interval?

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Here z* is 2.576 for 99%, z* is 1.96 for 0.95

Students abroad

70% of students from a sample size of 100 would like to Spend a semester abroad.

Build the 95% confidence interval and find the margin of error.

Repeat for sample size of 400 and compare

$$\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Students abroad

You should find that as we increase the sample size,
The interval gets smaller and so does the margin of error.

Large sample sizes are associated to narrower confidence intervals.

$$\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Sample size

The larger the better

Practical choices : it will depend on the margin of error.
If we set the margin of error E,
and have an estimate for p we can find n:

$$E = z^* \sqrt{\frac{p(1-p)}{n}}$$

$$E^2 = (z^*)^2 \frac{p(1-p)}{n}$$

$$n = (z^*)^2 \frac{p(1-p)}{E^2}$$

E is set,
z* is set,
p and (1-p) are set

Need to find n

Showerheads

The city passed an ordinance requiring residents to install low-flow showerheads. Water use has stayed the same. The city suspects that some homes have not complied with the law.
We need to estimate within 3% how many people complied with 95% confidence.
What sampling size should we use?

Use as a guess p =0.5

Showerheads

The city passed an ordinance requiring residents to install low-flow showerheads. Water use has stayed the same. The city suspects that some homes have not complied with the law.

We need to estimate within 3% how many people complied with 95% confidence

What sampling size should we use?

$$n = z^{*}\frac{p(1-p)}{E^{2}} = 1.96^{2}\frac{0.5*0.5}{0.03^{2}} = 1067.111 = 1068$$

Homework

Page 377
E1, E3, E5, E7, E9, E10, E11,E12, E13, E14,
E15, E16, E17, E18, E20,