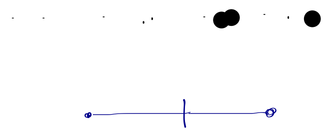
Welcome Back

What the heck were we doing in AP Stats before the break?

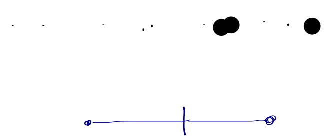
What the heck were we doing in AP Stats before the break?

Learning about making and Interpreting Confidence Intervals to estimate the "truth" about a parameter.

From every sample you can make a Confidence interval (CI)



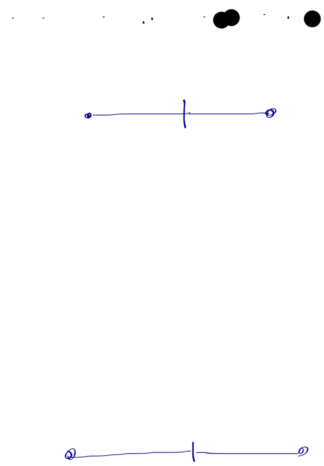
From every sample you can make a Confidence interval (CI) that has a set confidence level



From every sample you can make a Confidence interval (CI) that has a set confidence level

if you want to increase your confidence in capturing the true ___

it will....

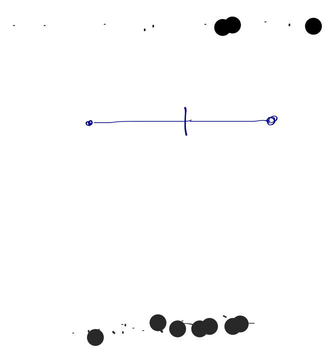


From every sample you can make a Confidence interval (CI)

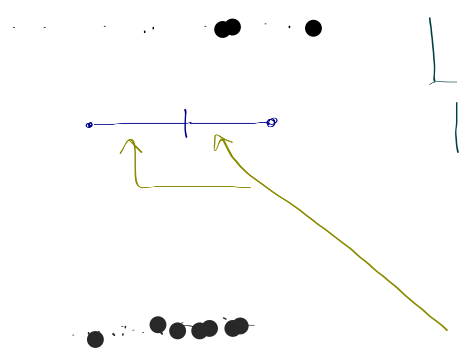
that has a set confidence level

if you want to increase your confidence in capturing the true ___

it will widen the interral.

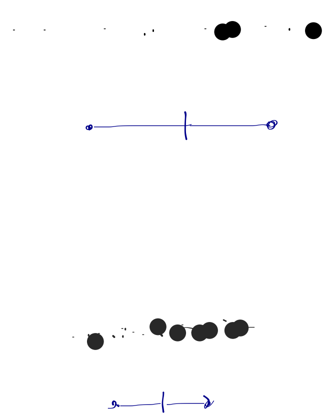


Larger samples have less variability on the average



Larger samples have less variability on the average

which will tend to Lower the margin of error



Larger samples have less variability on the average

which will tend to Lower the margin of error
which cause the confidence interval to narrow.

The tricky part is the interpretation of

- Confidence Interval (CI)
- Confidence Level

I may have led us to a slightly abbreviated version that could cause you to miss an important concept.

I may have led us to a slightly more abbreviated version

Important ideas:
Interpret Confid. Level

If we make many many 95 % confid. intervals, we expect about 95 % to capture the true parameter

I may have led us to a slightly more abbreviated version

Important ideas:

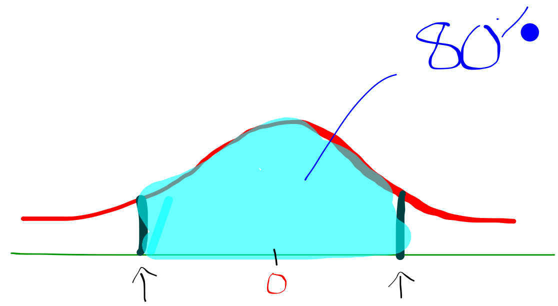
Interpret Confid. Level

If we make many many 95 % confid. intervals, we expect about 95 % to capture the true parameter.

If we take many many random samples of ______ and we construct a __% confidence interval from each one, then we expect...

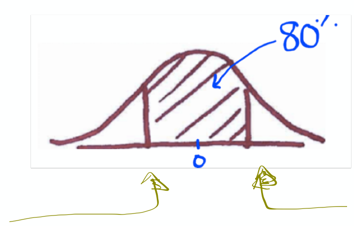
So look at the solutions to the 8.1 LCQ to help you refresh your mind what you learned before the break.

Pick Up the Warm Up



1) Use Table A, not technology, to calculate the two Z-scores that correspond with the diagram. (a portion of Table is shown below).

2) Now use technology to verify the corresponding Z-scores.



for z is the probability lying below z.

Table A *(Continued)* Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

between 1.28 and 1.29

≈ 1.28

Reminder: Terminology has changed

$\sigma_{\bar{x}}$ $\sigma_{\hat{p}}$ σ

standard error

Generic Formula to Construct a Confidence Interval:

$$\mathit{Statistic} \pm \underbrace{(\mathit{critical\ value}) \cdot (\mathit{st.}\ \cancel{\mathit{deviation}}\ \mathit{of\ sample\ statistic})}_{\text{Margin of Error (MOE)}}$$

Story of Inference

Ch. 2 → 68%/95%/99.7% rule for Normal Distrib.

95% are w/in 2 std. dev. of the mean

Ch. 2 → 68% / 95% / 99.7% rule for Normal Distrib.

95% are w/in 2 std dev. of the mean

Ch. 4 Purpose of random sample was to generalize to a larger population

A condition for Inference in ch 8 – 12

Ch. 2 → 68% / 95% / 99.7% rule for Normal Distrib.

95% are w/in 2 std dev. of the mean

Ch. 4 Purpose of random sample was to generalize to a larger population

A condition for Inference in ch 8 – 12

Ch 6 10% condition in the binomial setting to be sure there is independe. between trials.

Ch. 2 → 68%/95%/99.7% rule for Normal Distrib.

95% are w/in 2 std. dev. of the mean

Ch. 4 Purpose of random sample was to generalize to a larger population.

↑ A condition for Inference in ch. 8 – 12

Ch. 6 10% condition in the binomial setting to be sure there is independe. between trials.

Ch. 7 Used 10% condition to be sure certain formulas can be used for std. dev of sampling distributions.

Ch. 2 → 68%/95%/99.7% rule for Normal Distrib.

95% are w/in 2 std. dev. of the mean

Ch. 4 Purpose of random sample was to generalize to a larger population.

↑ A condition for Inference in ch. 8 – 12

Ch. 6 10% condition in the binomial setting to be sure there is independe. between trials.

Large Counts Condition in a binomial setting so it can be approximated by a Normal Distribution

Ch. 7 Used 10% condition to be sure certain formulas can be used for std. dev of sampling distributions.

Extended this reasoning to apply to a sampling distrib. of a proportion

Learning Target

State the conditions in which you can construct a confidence interval for a <u>population proportion.</u>

Today:

Tight for time, but there are Hershey's Kisses!

Use your formula sheet or notes to assist.

Lesson 8.2: Day 1:

Which way will the Hershey Kiss land?

When you toss a Hershey Kiss, it sometimes lands flat and sometimes lands on its side. What proportion of tosses will land flat? (We don't know the true proportion)

Each group of three (or four) selects a random sample of 50 Hershey's Kisses to bring back to their desks. Toss the 50 Kisses and then calculate the proportion that land flat. Let $\hat{p}$ = the proportion of the Kisses that land flat.

1. What is your **point estimate** for the true proportion that land flat? $\hat{p} =$ ______
2. Identify the population, parameter, sample and statistic.

   Population: ______________________ Parameter: ______________________

   Sample: ______________________ Statistic: ______________________

3. Was the sample a random sample? Why is this important?

Lesson 8.2: Day 1:

Which way will the Hershey Kiss land?

When you toss a Hershey Kiss, it sometimes lands flat and sometimes lands on its side. What proportion of tosses will land flat? (We don't know the true proportion)

Each group of three (or four) selects a random sample of 50 Hershey's Kisses to bring back to their desks. Toss the 50 Kisses and then calculate the proportion that land flat. Let $\hat{p}$ = the proportion of the Kisses that land flat.

1. What is your **point estimate** for the true proportion that land flat? ________

2. Identify the population, parameter, sample and statistic.

   Population: All Hershey's Kisses Parameter: p = true proportion that land flat

   Sample: 50 Hershey's Kisses Statistic: $\hat{p}$ = ________ yours

3. Was the sample a random sample? Why is this important?

   Yes. This is vital so we can generalize (make inferences) to the whole population.

---

4. What is the formula for calculating the standard deviation of the sampling distribution of $\hat{p}$? p

5. What condition must be met to use this formula? Has it been met?

4. What is the formula for calculating the standard deviation of the sampling distribution of $\hat{p}$?

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

5. What condition must be met to use this formula? Has it been met?

10% Condition $n < \frac{1}{10}N$

$50 < \frac{1}{10}$ of all Kisses

6. We don't know the value of p (that's the whole point of a confidence interval) so we will use $\hat{p}$ instead. Calculate the standard deviation.

7. Would it be appropriate to use a normal distribution to model the sampling distribution of $\hat{p}$? Justify your answer.

6. We don't know the value of p (that's the whole point of a confidence interval) so we will use $\hat{p}$ instead. Calculate the standard deviation.

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{.42(.58)}{50}} = .0698$$

7. Would it be appropriate to use a normal distribution to model the sampling distribution of $\hat{p}$? Justify your answer.

Large Counts

$50(.42) = 21 \geq 10$

$50(.58) = 29 \geq 10$

np

$n(1-p)$

So, yes a normal distrib is appropriate

6. We don't know the value of p (that's the whole point of a confidence interval) so we will use $\hat{p}$ instead. Calculate the standard deviation.

Standard Error

$\cancel{SE_{\hat{p}}}$

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{.42(.58)}{50}} = .0698$$

7. Would it be appropriate to use a normal distribution to model the sampling distribution of $\hat{p}$? Justify your answer.

Large Counts

$50(.42) = 21 \geq 10$

$50(.58) = 29 \geq 10$

So, yes a normal distrib is appropriate

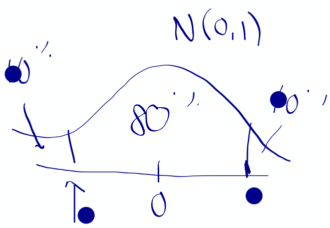
8. In a normal distribution, 95% of the data lies within 2 standard deviations of the mean. This value is called the **critical value.** Use table A or $invNorm$ to find these critical values:

80% of the data lies within ________ standard deviations of the mean

90% of the data lies within ________ standard deviations of the mean

95% of the data lies within ________ standard deviations of the mean

99% of the data lies within ________ standard deviations of the mean



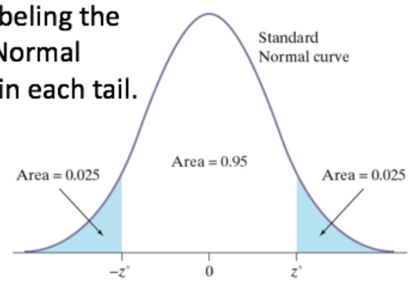
Constructing a Confidence Interval for *p*

How do we get the critical value z* for our confidence interval?

Constructing a Confidence Interval for *p*

How do we get the critical value z* for our confidence interval?

Finding the critical value z* for a 95% confidence interval starts by labeling the middle 95% under a standard Normal curve and calculating the area in each tail.

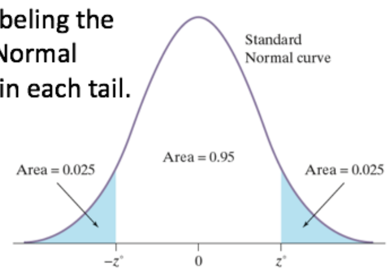


Constructing a Confidence Interval for *p*

How do we get the critical value z* for our confidence interval?

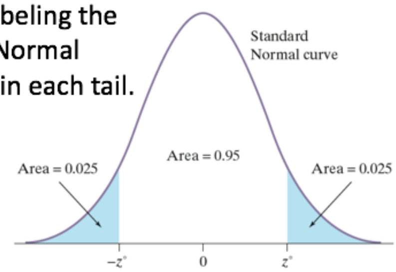
Finding the critical value z* for a 95% confidence interval starts by labeling the middle 95% under a standard Normal curve and calculating the area in each tail.

Using Table A: Search the body of Table A to find the point $-z^*$ with area 0.025 to its left. The entry $z = -1.96$ is what we are looking for, so $z^* = 1.96$.



How do we get the critical value z* for our confidence interval?

Finding the critical value z* for a 95% confidence interval starts by labeling the middle 95% under a standard Normal curve and calculating the area in each tail.



Using Table A: Search the body of Table A to find the point $-z^*$ with area 0.025 to its left. The entry z = – 1.96 is what we are looking for, so z^* = 1.96.

Using technology: The command invNorm(area:0.025, mean:0, SD:1) gives z = –1.960, so z^* = 1.960.

8. In a normal distribution, 95% of the data lies within ______ standard deviations of the mean This value is called the **critical value.** Use table A or $invNorm$ to find these critical values:

80% of the data lies within 1.282 standard deviations of the mean

90% of the data lies within 1.645 standard deviations of the mean

95% of the data lies within 1.960 standard deviations of the mean

99% of the data lies within 2.576 standard deviations of the mean

9. Calculate the **margin of error** for a 95% interval by multiplying the critical value and standard deviation you found. Show your work.

10. Find the 95% confidence interval using **point estimate +/- margin of error**.

9. Calculate the **margin of error** for a 95% interval by multiplying the critical value and standard deviation you found. Show your work.

example

critical value

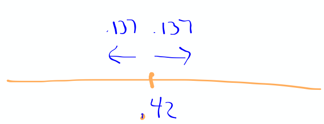
$1.96 \times 0.0698 = 0.137$

std dev.

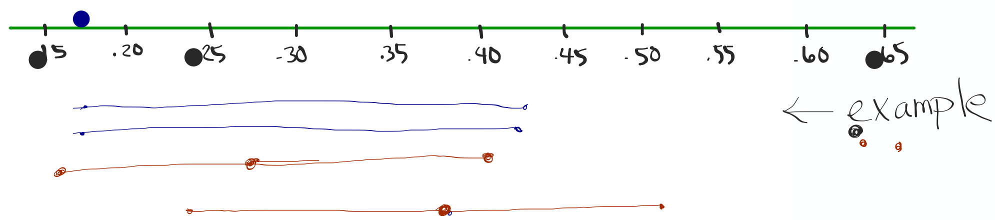
10. Find the 95% confidence interval using **point estimate +/- margin of error**.

$.42 \pm .137 = (.283, .557)$

PE



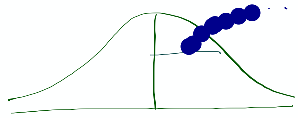
11. Add your interval to the graph on the board. Sketch the graph below.



12. What do you think is the true proportion of kisses that land flat is?

Back side

Sampling Distribution of proportions ($\hat{p}$)



Standard Error

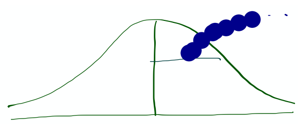
Values of $\hat{p}$

p unknown

$$\cancel{SE}_{\hat{p}} \quad S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}$ because we don't know value of p

Sampling Distribution of proportions ($\hat{p}$)



Standard Error

Values of $\hat{p}$

p unknown

$$\cancel{SE}_{\hat{p}} \quad S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}$ because we don't know value of p

Just look

Like the standard deviation, the standard error describes how much the the sample proportion, $\hat{p}$, typically varies from the pop. proportion in repeated random samples of size n

Constructing a Confidence Interval for *p*

There are three conditions that must be met for this formula to be valid—one for each of the three components in the formula.

$$\underset{1}{\hat{p}} \pm \underset{2}{z^*} \underset{3}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

One-Sample z-Interval For a Population Proportion

When conditions are met, a $C\%$ confidence interval for an unknown proportion, p, is

Random Condition
(from SRS or from Experiment with rand. Assignment)

Large Counts Condition
So, we can assume approx. Norm. Dist.
$np \geq 10$ *and* $n(1-p) \geq 10$

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

the 10% Condition
to check that individual observations are independent.

unless you are told there is replacement

One-Sample z-Interval For a Population Proportion

When conditions are met, a $C\%$ confidence interval for an unknown proportion, p, is

Random Condition
(from SRS or from
Experiment with rand. Assignment)

Large Counts Condition
So, we can assume approx. Norm. Dist.
$np \geq 10$ and $n(1-p) \geq 10$

the 10% Condition
to check that individual observations are independent.

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

Summary for Constructing a Confidence Interval for *p* (proportions)

Important ideas:

CONDITIONS	Critical Values	Confidence Intervals for p

Summary for Constructing a Confidence Interval for *p* (proportions)

Important ideas: CONDITIONS	Critical Values	Confidence Intervals for p
① Random ② 10% Condition $n < \frac{1}{10}N$ ③ Normal/Large Counts $n\hat{p} \geq 10$, $n(1-\hat{p}) \geq 10$		

Each group of three (or four) selects a random sample of 50 Hershey's Kisses to bring back to their desks. Toss the 50 Kisses and then calculate the proportion that land flat. Let $\hat{p}$ = the proportion of the Kisses that land flat.

1. What is your **point estimate** for the true proportion that land flat? ________

2. Identify the population, parameter, sample and statistic.

   Population: All Hershey's Kisses Parameter: p = true proportion that land flat

   Sample: 50 Hershey's Kisses Statistic: $\hat{p}$ =

   yours

Condition #1

3. Was the sample a random sample? Why is this important?

   Yes. This is vital so we can generalize (make inferences) to the whole population.

4. What is the formula for calculating the standard deviation of the sampling distribution of $\hat{p}$?

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

Condition #2

5. What condition must be met to use this formula? Has it been met?

10% Condition $\quad n < \frac{1}{10}N$

$50 < \frac{1}{10}$ of all Kisses.

6. We don't know the value of p (that's the whole point of a confidence interval) so we will use $\hat{p}$ instead. Calculate the standard deviation.

Standard Error

~~$SE_{\hat{p}}$~~

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{.42(.58)}{50}} = .0698$$

7. Would it be appropriate to use a normal distribution to model the sampling distribution of $\hat{p}$? Justify your answer.

Condition #3

Large Counts

$50(.42) = 21 \geq 10$

$50(.58) = 29 \geq 10$

So, yes a normal distrib is appropriate

Summary for Constructing a Confidence Interval for *p* (proportions)

Important ideas:

CONDITIONS

1. Random
2. 10% Condition
   $n < \frac{1}{10}N$
3. Normal/Large Counts
   $n\hat{p} \geq 10$
   $n(1-\hat{p}) \geq 10$

Critical Values

90% $z^* = 1.645$

95% $z^* = 1.960$

99% $z^* = 2.576$

Confidence Intervals for p

Summary for Constructing a Confidence Interval for *p* (proportions)

Important ideas:

CONDITIONS

1. Random
2. 10% Condition
   $n < \frac{1}{10}N$
3. Normal/Large Counts
   $n\hat{p} \geq 10$
   $n(1-\hat{p}) \geq 10$

Critical Values

90% $z^* = 1.645$

95% $z^* = 1.960$

99% $z^* = 2.576$

To find any %
use
invNorm(tail)

$\text{invNorm}\left(\frac{1-\%}{2}\right)$

Confidence Intervals for p

Summary for Constructing a Confidence Interval for *p* (proportions)

Important ideas:

CONDITIONS	Critical Values	Confidence Intervals for p
① Random	90% $z^* = 1.645$	Point Estim $\pm z^* \underbrace{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$
② 10% Condition $n < \frac{1}{10}N$	95% $z^* = 1.960$	Standard error
③ Normal/Large Counts $n\hat{p} \geq 10$ $n(1-\hat{p}) \geq 10$	99% $z^* = 2.576$	$S_{\hat{p}}$
	To find any % use invNorm(tail)	
	invNorm$\left(\frac{1-\%}{2}\right)$	

bottom of front side

III. Sampling Distributions and Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$

Confidence interval: statistic $\pm$ (critical value)(standard error of statistic)

Chi-square statistic: $\chi^2 = \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

back side

III. Sampling Distributions and Inferential Statistics *(continued)*

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Later

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Sta San
For one population: $\overline{X}$	$\mu_{\overline{X}} = \mu$	$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$	
For two populations: $\overline{X}_1 - \overline{X}_2$	$\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$	$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\overline{X}_1}$

Sampling distributions for simple linear regression:

Sleep
Awareness

Sleep Awareness Week begins in the spring with the release of the National Sleep Foundation's annual poll of U.S. sleep habits and ends with the beginning of daylight savings time, when most people lose an hour of sleep. In the foundation's random sample of 1029 U.S. adults, 48% reported that they "often or always" got enough sleep during the last 7 nights.

1. Identify the parameter of interest.

2. Check if the conditions for constructing a confidence interval for p are met.

Sleep Awareness Week begins in the spring with the release of the National Sleep Foundation's annual poll of U.S. sleep habits and ends with the beginning of daylight savings time, when most people lose an hour of sleep. In the foundation's random sample of 1029 U.S. adults, 48% reported that they "often or always" got enough sleep during the last 7 nights.

1. Identify the parameter of interest.

p = true prop. of all US adults who "often or always" got enough sleep during the last 7 days.

2. Check if the conditions for constructing a confidence interval for *p* are met.

(1) Random
(2) 10%
(3) Normal/Large Counts

Sleep Awareness Week begins in the spring with the release of the National Sleep Foundation's annual poll of U.S. sleep habits and ends with the beginning of daylight savings time, when most people lose an hour of sleep. In the foundation's random sample of 1029 U.S. adults, 48% reported that they "often or always" got enough sleep during the last 7 nights.

1. Identify the parameter of interest.

p = true prop. of all US adults who "often or always" got enough sleep during the last 7 days.

2. Check if the conditions for constructing a confidence interval for *p* are met.

(1) Random : a random sample of US Adults ✓
(2) 10% $1029 < \frac{1}{10}$(All US adults) ✓
(3) Normal/Large Counts $1029(.48) = 493.92 \geq 10$ ✓
$1029(.52) = 535.08 \geq 10$

3. Find the critical value for a 98% confidence interval. Then calculate the interval.

$z^* = \text{invNorm}(.01) = 2.236$

Tail Proportion

$\frac{1-.98}{2} = .01$

4. Interpret the interval in context.

3. Find the critical value for a 98% confidence interval. Then calculate the interval.

$z^* = \text{invNorm}(.01) = 2.236$

Tail Proportion

$\frac{1-.98}{2} = .01$

$\hat{p} \pm z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow .48 \pm 2.326\sqrt{\frac{.48(.52)}{1029}}$

4. Interpret the interval in context.

3. Find the critical value for a 98% confidence interval. Then calculate the interval.

Tail Proportion

$\frac{1-.98}{2} = .01$

$z^* = \text{invNorm}(.01) = 2.236$

$\hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow .48 \pm 2.326\sqrt{\frac{.48(.52)}{1029}}$

$.48 \pm .036 = (.444, .516)$

4. Interpret the interval in context.

We are 95% confident that the interval from .444 to .516 captures the true proportion of all US adults who report that they often or always got enough sleep in last 7 days.

See your Unit 2 PPC- FRQ

8.2.....29, 31, 35, 35, 37, 39, <u>59</u>

and study pp. 510-516

Two LCQ's have been dropped in Synergy