Name_____

AP Statistics CED 7.1 Daily Video 1 (Skill 1.A)

Introducing Statistics: Why Should I Worry About Error?

What Will We Learn?

How do we identify evidence for a claim?

How do we determine if the evidence for a claim is convincing?

Bonus vs. Rebate

Are people more likely to spend money if it is called a bonus or if it is called a rebate? Volunteer college students were given $50 with no strings attached. 25 of the 47 students randomly assigned to be told that the money was "bonus income," while the other 22 students were told the money was a "tuition rebate." After one week, the students were asked to report how much of the $50 they spent. Here are the data, along with the mean amount spent by each group.

Amount Spent in Dollars (Bonus Income)				
0	0	0	0	0
0	0	0	0	10
10	10	20	25	26
30	30	40	50	50
50	50	50	50	50

$\bar{x}_{BONUS}$ = $22.04

Amount Spent in Dollars (Tuition Rebate)				
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	10	20	30	50
50	50			

$\bar{x}_{REBATE}$ = $9.55

Bonus vs. Rebate

Do these data provide convincing evidence that the bonus wording causes college students like the _____ in this study to spend more money, on average, than the tuition rebate wording?

What is the _____ that more money would be spent using the _____ income wording? If the wording made _____ in the amount spent, the mean amount spent by _____ group should be _____. Therefore, we expect a _____ in means of _____.

In the study, $\bar{x}_{BONUS} - \bar{x}_{REBATE}$ = _____ = _____ > _____

Bonus vs. Rebate

What are the _____ explanations for the _____ found in this study?

1. The wording made _____ in the amount spent, and the $12.49 difference in _____ amount spent occurred due to _____ in the _____ to groups.
2. The bonus wording _____ students to spend more, on _____ than the rebate wording.

To be convinced that Explanation #2 is correct, we need to know _____ it would be to get a _____ in means of _____ by chance alone. If it is unlikely to get _____ by _____ in the _____ assignment, we can rule out Explanation #1 and _____ that Explanation #2 is correct.

Bonus vs. Rebate

To estimate $P(\bar{x}_{BONUS} - \bar{x}_{REBATE} \geq \$12.49)$ when _____ makes _____ in the amount that would be spent, we can perform a _____. Because we are _____ that wording makes _____, subjects are expected to spend the _____ amount no matter which group they're in.

So, we can _____ assign the subjects to groups (_____), find the _____ amount for each group and calculate the _____ in means.

Name_____

Bonus vs. Rebate - Simulation

Here is one possible random assignment:

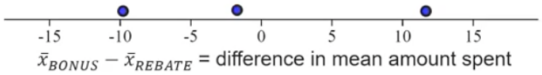
Amount Spent in Dollars (Bonus Income)				
0	0	20	0	50
0	0	50	0	0
10	0	50	20	0
0	0	10	0	50
0	30	0	0	0

$\bar{x}_{BONUS}$ = $11.60

Amount Spent in Dollars (Tuition Rebate)				
50	50	50	50	30
0	0	10	40	0
0	0	26	0	10
50	30	0	0	50
0	25			

$\bar{x}_{REBATE}$ = $21.41

In this instance $\bar{x}_{BONUS} - \bar{x}_{REBATE}$ = -$9.81



Below are the results of 1000 trials of this simulation. Based on the simulation, would it be surprising to get a difference in means of $\bar{x}_{BONUS} - \bar{x}_{REBATE}$ = $12.40 or greater just by chance variation in the random assignment?

Amount Spent in Dollars (Bonus Income)				
10	50	50	50	0
0	0	0	50	50
50	25	0	40	0
0	50	0	20	26
20	0	50	0	0

$\bar{x}_{BONUS}$ = $21.64

Amount Spent in Dollars (Tuition Rebate)				
0	10	0	0	0
10	0	0	50	0
30	30	0	30	10
0	0	0	0	50
0	0			

$\bar{x}_{REBATE}$ = $10.00

In this instance $\bar{x}_{BONUS} - \bar{x}_{REBATE}$ = _____

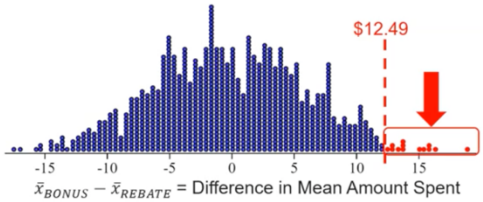
Amount Spent in Dollars (Bonus Income)				
0	0	10	30	0
0	0	0	0	0
50	50	20	0	0
0	0	50	0	50
50	0	30	0	50

$\bar{x}_{BONUS}$ = $15.60

Amount Spent in Dollars (Tuition Rebate)				
10	0	0	10	0
0	0	10	50	26
0	50	0	25	50
50	30	0	20	0
0	40			

$\bar{x}_{REBATE}$ = $16.86

In this instance $\bar{x}_{BONUS} - \bar{x}_{REBATE}$ = _____



Amount Spent in Dollars (Bonus Income)				
0	0	50	0	0
25	50	0	0	0
50	30	50	26	10
50	50	0	0	30
0	10	0	0	20

$\bar{x}_{BONUS}$ = $18.04

Amount Spent in Dollars (Tuition Rebate)				
50	0	0	0	50
10	0	0	0	40
20	0	0	10	0
50	30	0	50	0
0	0			

$\bar{x}_{REBATE}$ = $14.09

In this instance $\bar{x}_{BONUS} - \bar{x}_{REBATE}$ = _____.
(Plot this point on the graph above.)

_____. In ____ out of _____ trials,

$\bar{x}_{BONUS} - \bar{x}_{REBATE}$ was _____.

Thus, $P(\bar{x}_{BONUS} - \bar{x}_{REBATE}$ = $12.49) ≈ ____=____

Bonus vs. Rebate

Do these data provide _____ that the bonus wording causes college students like the ones in this study to spend more money, on average, than the tuition rebate wording?

1. The wording made _____ in the amount spend, and the $12.49 difference in mean amount spent occurred due to _____ in the random assignment groups.
2. The bonus wording _____ students to spend more, on average, than the rebate wording.

Because it is _____ (*p*-value ≈ _____) to get a difference in means of $\bar{x}_{BONUS} - \bar{x}_{REBATE}$ = $12.49 or _____ by chance alone when _____ in the amount that would be spent, we can rule out Explanation _____. These data ____ provide _____ that the bonus wording causes college students _____ in this study to spend more money, on average than the tuition rebate wording.

What Should We Take Away?

How do we identify evidence for a claim?

Show that the _____ of a study are _____ with the claim.

How do we determine if the evidence for a claim is convincing?

- **Consider the _____ explanations for the evidence (_____ chance, _____ effect).**
- **_____ the probability of getting evidence as _____ than the observed evidence by _____ alone.**
- **If you can eliminate random chance as a _____ explanation for the evidence, the evidence for the claim is convincing.**

Name_____

AP Statistics CED 7.2 Daily Video 1 (Skill 3.C)

Constructing a Confidence Interval for Population Mean

What Will We Learn?

Why is a *z* critical value not appropriate for creating a confidence interval for a population mean?

What critical value is appropriate for creating a confidence interval for a population mean?

Confidence Intervals: Structure

From Unit 6 videos,

CI = _____ ± _____

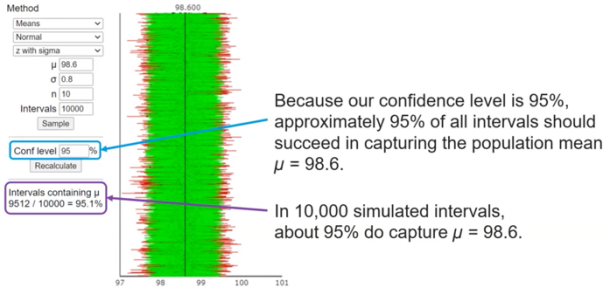
CI = _____ ± (_____)(_____)

Expected formula to estimate a population mean: CI = $\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$

Human Body Temperature

Suppose the human body temperatures are approximately normally distributed with a μ = 98.6°F and a standard deviation of σ = 0.08°F. Assuming conditions have been met, will the formula CI = $\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$ work as expected? To find out, let's us technology to simulate 95% confidence intervals for μ based on random samples of *n* = 10 people.

z* with σ



But this is unrealistic because we assume that σ is known. In situations where we must truly _____ the population mean μ, its unreasonable to think we would know the population standard deviation σ.

How should we proceed? We can use the sample standard deviation *s* as an estimate of σ.

Confidence Intervals: Structure

To estimate a population mean:

CI = $\bar{x} \pm z\frac{s}{\sqrt{n}}$

Under the right conditions, the sampling distribution of $\bar{x}$ is approximately _____.

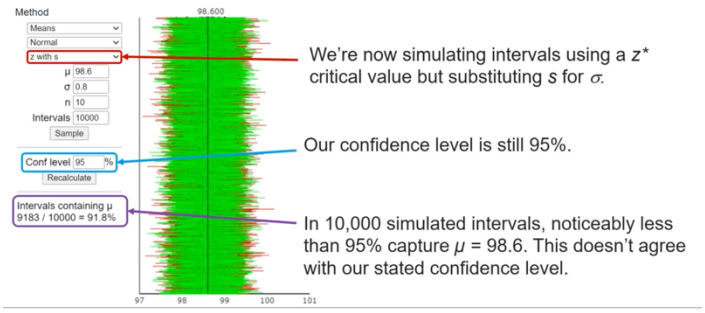
For random samples, $\bar{x}$ is an _____estimate of μ.

Using *s* as an estimate of σ seems like it should do the trick. Let's simulate confidence intervals for human body temperatures to see how it performs.

Name_____

z^* with s

We see that now only about _____ of the simulated intervals are capturing _____.



Houston, We *Still* Have a Problem

Using _____ as an estimate of _____ causes the _____ that actually succeed in capturing ____ to be less than our _____ confidence level.

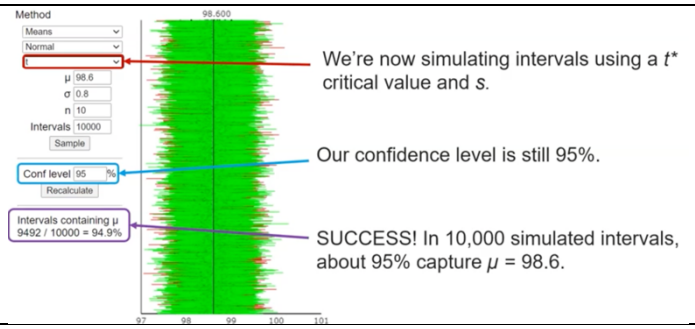
How do we fix this?

We use a different critical value, t^*. The critical value of _____ come from a ____ distribution, not a _____ distribution.

To estimate the population mean: CI = $\bar{x} \pm t * \frac{s}{\sqrt{n}}$

t^* with s

Now we see that, in fact, about _____ of our intervals _____ capture μ, our population mean.



Time for a Recap

In upcoming videos, we will learn how to _____ and _____ confidence intervals for a population mean _____.

In practice, the population _____ is almost never known, so we use the sample standard deviation s to _____ it.

Using a _____ critical value with ____ produces intervals that capture the population mean _____ often than the advertised confidence level, so we use a _____ critical value instead. You can think of the _____ critical value as an adjustment to the _____ critical value that makes the confidence intervals slightly _____ so that the percentage of intervals that do capture _____ is equal to the confidence level.

What Should We Take Away?

Why is a z critical value not appropriate for creating a confidence interval for a population mean?

A z critical value is appropriate when we know _____. This _____ occurs.

What critical value is appropriate for creating a confidence interval for a population mean?

A _____ critical value.

Name_____

AP Statistics CED 7.2 Daily Video 2 (Skill 4.C)

Constructing a Confidence Interval for a Population Mean

What Will We Learn?

How do we identify an appropriate confidence interval procedure for a population mean?
How do we verify the conditions for calculating a confidence interval for a population mean?

Some Reminders…

There are two major types of statistical inference in Units 6 – 9:

Confidence Intervals

- used to _____ the value of a _____.
- an _____ - rather than a single _____ - used to estimate a _____ to account for _____ variability

Significance Tests

- used to _____ about the value of a population parameter
- assess whether the _____ supporting a claim is likely or unlikely to happen by _____.

The Structure of Unit 7

In Unit 7, we will focus on _____ data that can be summarized by calculating the _____ in a sample or treatment group.

Topics 7.2 – 7.3 → _____ for a population _____
Topics 7.4 – 7.5 → _____ for a population _____
Topics 7.6 – 7.7 → _____ for a difference in _____
Topics 7.8 – 7.9 → _____ for a difference in _____

Sweet Sugary Goodness

A bakery owner purchases powdered sugar, also called confectioners' sugar or 10X sugar, from a new food wholesaler. Because the food wholesaler is new, the bakery owner wants to make sure that the bags were properly filled, on average. Here are the weights (in grams) of 10 randomly selected bags from the food wholesaler. Calculate and interpret a 95% confidence interval for the mean weight μ of all bags filled by this wholesaler.

910	919	900	913	904
913	903	914	893	899

Identifying the Procedure

You have already learned about a confidence interval for a population proportion. But, in this case we are estimating a population _____.

When the goal is to estimate the _____ in a _____ population, we use a _____.

Name_____

Checking the Conditions

Here are the conditions for a _____ for a population _____.

To check for _____:

1. The data are collected using a _____ from the population or a _____.
2. When sampling _____ replacement, the _____ is less than or equal to _____ of the population size.

To check that the shape of the _____ is approximately normal:

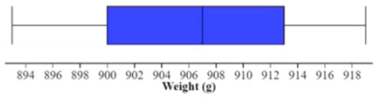
3. The sample size is _____ ($n \geq$ _____)
   OR
   If $n <$ ___, the sample data are free from strong _____ or _____.

Checking the Conditions. (Be sure to place a ü after each condition!)

A bakery owner wants to make sure that the bags of powdered sugar were properly filled, on average. Here are the weights (in grams) of 10 randomly selected bags from the good wholesaler. Check if the conditions for calculating a confidence interval are met.

910	919	900	913	904
913	903	914	893	899

1. The 10 bags were _____ selected. ü
2. It's reasonable to believe 10 bags is _____ to _____ of all bags from a food wholesaler.
3. A boxplot of the weights shows no strong skewness or outliers. (Use technology to create boxplot and sketch.)



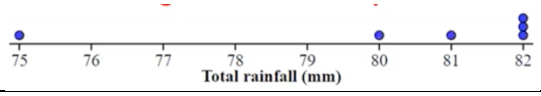
Avril Shower…

Avril wants to estimate the average amount of rainfall for all houses in her county during the month of April. She gets 6 classmates to put rain gauges in their yards and record the total rainfall during the month of April. Here are their data, in millimeters. Check if the conditions for calculating a confidence interval are met.

82	80	81	82	75	82

1. The 6 houses were _____ randomly selected. û
2. It's reasonable to believe 6 houses is less than or equal to _____ of all houses in a county.
3. A dotplot of the total rainfall shows _____ and a _____.

The conditions are _____ met.



What Should We Take Away?

How do we identify an appropriate confidence interval procedure for a population mean?

A _____- sample ___ interval for a population _____.

How do we verify the conditions for calculating a confidence interval for a population mean?

1. **The data are collected using a random sample from the population or a randomized experiment.**
2. **When sampling without replacement, the _____ is less than or equal to _____ of the population size.**
3. **The sample size is _____ ($n \geq$ _____). OR**
4. **If $n <$ ___, the sample data are free from strong _____ or _____.**

Name_____

AP Statistics CED 7.2 Daily Video 3 (Skill 3.D)

Constructing a Confidence Interval for a Population Mean

What Will We Learn?

How do we determine the margin of error when estimating a population mean?
How do we calculate a confidence interval for a population mean?

Sweet Sugary Goodness

A bakery owner purchases powdered sugar, also called confectioners' sugar or 10X sugar, from a new food wholesaler. Because the food wholesaler is new, the bakery owner wants to make sure that the bags were properly filled, on average. Here are the weights (in grams) of 10 randomly selected bags from the food wholesaler. Calculate a 95% confidence interval for the mean weight μ of all bags filled by this wholesaler.

910	919	900	913	904
913	903	914	893	899

Calculating the Margin of Error

margin of error = (critical value)(_____)

The _____ of a statistic is an estimate of the _____ of the _____ of the statistic. From Topic 5.7, the standard deviation of the sampling distribution of $\bar{X}$ is: $\sigma_x = \frac{\sigma}{\sqrt{n}}$. Because we don't know the value of σ, we replace it with s to get the standard error of $\bar{X}$: $SE_{\bar{x}} = \frac{s}{\sqrt{n}}$

Calculating the Margin of Error

margin of error = (_____)(standard error of statistic)

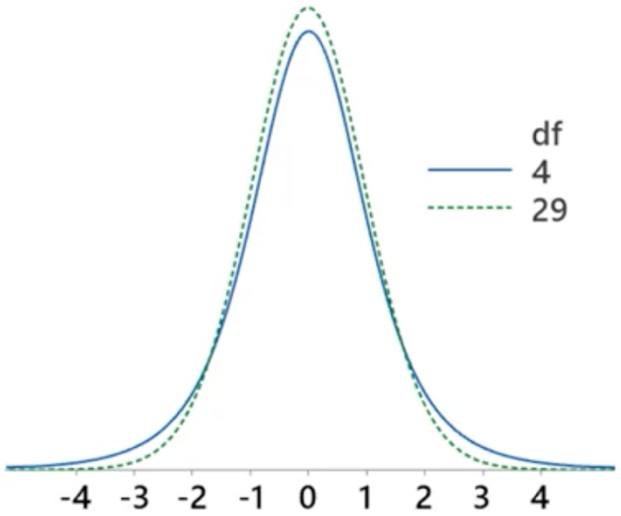
The _____ is a multiplier that makes the margin of error large enough to give a specific amount of _____ that the interval contains the value being estimated. As we saw in Topic 7.1 video, when estimating a population mean, the critical value comes from a _____.

***t* Distributions**

The mean of a *t* distribution is _____.
The variability of a _____ is determined by it _____.

For estimating a population mean, df = _____.
The shape of a *t* distribution is like a standard normal distribution, but with _____ and a slightly lower _____.

As degrees of freedom increase, the tails get _____ and the peak gets _____.

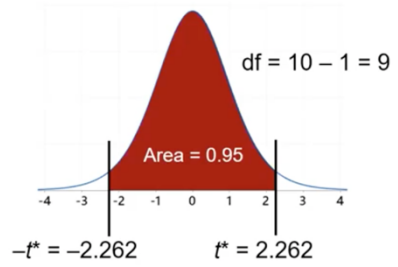


Name_____

Calculating the Margin of Error

Margin of error = (critical value)(standard error of statistic)

To find the critical value _____ for a 95% confidence interval in the powdered sugar example, find the boundaries encompassing the _____ 95% of the ___ distribution with df = _____.



(Use appropriate technology or **Table B** to find the critical value

The critical value is t^* = _____.

Calculating the Confidence Interval

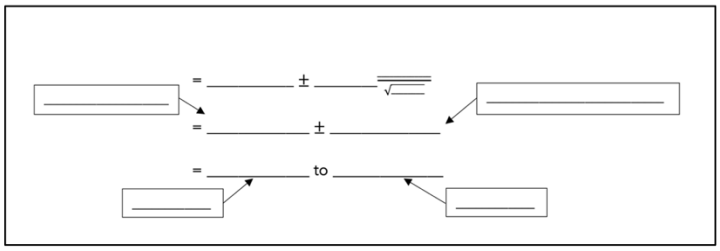
CI = _____ ± _____

CI = _____ ± (_____)(_____)

For our example about powdered sugar, we used technology to find that $\bar{x}$ = _____, s = _____ and n = _____.

Therefore, we can substitute in values to find:

$$CI = \bar{x} \pm t * \frac{s}{\sqrt{n}}$$



Fiddler Feeding Frenzy

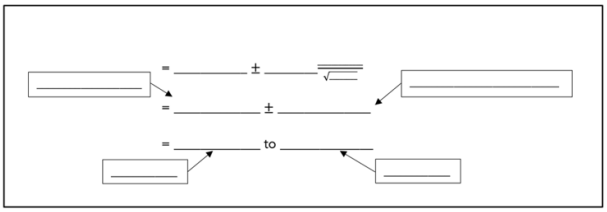
Fiddler crabs eat by sifting through sand to find food particles and then scooping them into their mouths. Wildlife biologists studied a random sample of one species of fiddler crabs to estimate the speed at which they feed, measured in number of scoops per 30 seconds. For the 40 crabs in the sample, $\bar{x}$= 67.65 scoops per 30 seconds and s = 6.61 scoops per 30 seconds. Calculate a 90% confidence interval for the mean feeding rate of all fiddler crabs of this species.

First, use technology of Table B to find the critical value of t^* _____.

For df=_____ and 90% confidence,

t^* = _____, $\bar{x}$ = _____, s = _____ and n = _____

$$CI = \bar{x} \pm t * \frac{s}{\sqrt{n}}$$



What Should We Take Away?

How do we determine the margin of error when estimating a population mean?

margin of error = (_____)(_____)

margin of error = $t * \frac{s}{\sqrt{n}}$ where t^* come from a _____ with _____ degrees of freedom.

How do we calculate a confidence interval for a population mean?

CI = _____ ± _____; thus: CI = $\bar{x} \pm t * \frac{s}{\sqrt{n}}$

Name_____

AP Statistics CED 7.3 Daily Video 1 (Skill 4.D)

Justifying a Claim About a Population Mean Based on a Confidence Interval

What Will We Learn?

How do we interpret a confidence interval for a population mean?
How do we justify a claim based on a confidence interval for a population mean?

Sweet Sugary Goodness

A bakery owner purchases powdered sugar, also called confectioners' sugar or 10X sugar, from a new food wholesaler. Because the food wholesaler is new, the bakery owner wants to make sure that the bags were properly filled, on average. Here are the weights (in grams) of 10 randomly selected bags from the food wholesaler. Calculate and interpret a 95% confidence interval for the mean weight μ of all bags filled by this wholesaler.

910	919	900	913	904
913	903	914	893	899

Interpreting the Confidence Interval

"We are C% confident that the interval from _____ to _____ captures the [parameter to be estimated]." From Topic 7.2, Video 3, the 95% confidence interval for the powdered sugar example is _____ to _____.

"We are _____ confident that the interval from _____ to _____ captures the _____."

Justifying a Claim

The manufacturer claims that the bags contain an average of 907 grams of powdered sugar. The bakery owner is worried that the bags weigh less than 907 grams, on average.

Is there convincing evidence that the bags of powdered sugar are being under-filled?

Because _____ grams is ___ the 95% confidence interval (900.92 gram to 912.68 grams), 907 grams is a _____ for ___, the mean weight of _____ bags filled by _____ wholesaler. Thus, there is _____ that the bags of powdered sugar are being under-filled.

Fiddler Feeding Frenzy

Wildlife biologists studied a random sample of 40 fiddler crabs to see how quickly they eat. The 90% confidence interval for the mean feeding rate of all fiddler crabs of this species is 67.65 ± 1.76 scoops per 30 seconds.

(a) interpret the confidence interval.
(b) Based on the interval, is there convincing evidence that the average feeding rate for this species is faster (greater) than 2 scoops per second?

Name_____

Fiddler Feeding Frenzy

(a) Interpret the confidence interval.

67.65 ± 1.76 → _____ to _____ scoops per 30 seconds

"We are _____ confident that the interval _____ captures the _____

(b) Based on the interval, is there convincing evidence that the average feeding rate for this species is faster (greater) than 2 scoops per second?

Convert 2 scoops per 1 second = _____

Because all values in the interval (_____) are greater than _____, there is _____ that the average feeding rate for _____ is faster (greater) than 2 scoops per second.

What Should We Take Away?

How do we interpret a confidence interval for a population mean?

"We are _____ confident that the interval from _____ to _____ captures the _____."

How do we justify a claim based on a confidence interval for a population mean?

- **If _____ the values in the confidence interval are _____ with the claim, there is _____ evidence for the claim.**
- **If _____ of the values in the confidence interval are _____ with the claim, there is _____ evidence for the claim.**

Name_____

AP Statistics CED 7.3 Daily Video 2 (Skill 4.A)

Justifying a Claim About a Population Mean Based on a Confidence Interval

What Will We Learn?

How do we interpret the confidence **level** for a confidence interval for a population mean? How do the sample size and confidence level affect the margin of error for a confident interval for a population mean?

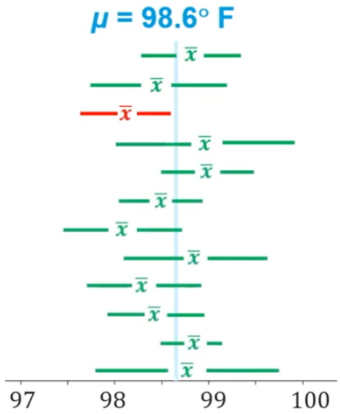
Human Body Temperature

Suppose the human body temperatures are approximately normally distributed with a μ = 98.6°F and a standard deviation of σ = 0.08°F. Suppose we select a random sample of n = 10 people, record the body temperature of each person and calculate a 95% confidence interval for the mean human body temperature of all people.

Human Body Temperature

Population: μ = 98.6° F; σ = 0.8° F

Shape: approximately normal



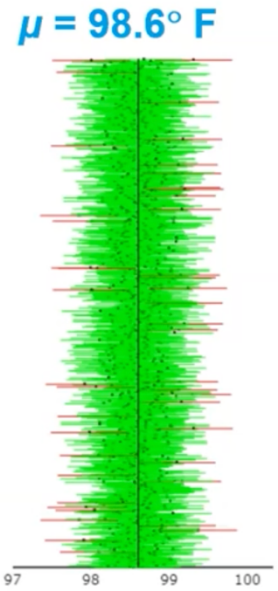
Some possible samples:

$\bar{x}$ = 98.79; 95% CI = _____ to _____

$\bar{x}$ = 98.46; 95% CI = _____ to _____

$\bar{x}$ = 98.10; 95% CI = _____ to _____

$\bar{x}$ = 98.89; 95% CI = _____ to _____



In the 12 samples to the right,

_____ of the intervals have capture μ = 98.6° F, the mean body temperature of all humans.

But if we simulate 1000 random samples we find:

In this simulation, _____ = _____ of the intervals have captured μ = 98.6° F, the mean body temperature of all humans.

Interpreting the Confidence Level

In general, here is how to interpret a confidence **level** for a population parameter:

"In repeated random sampling with the _____ sample size, approximately _____ of those _____ confidence intervals created will capture the population mean."

If we take _____ random samples of _____ **from the population of all humans** and use each sample to construct a _____ confidence interval **for the mean body temperature of _____ people**, about _____ of those intervals would capture the population mean.

Remember to write in context. Always.

Name_____

Factors That Affect the Margin of Error

Recall that confidence intervals in AP Statistics have the following structure:

CI = _____ ± _____

The _____ of the confidence interval is _____ the margin of error.

We generally prefer _____ confidence intervals (_____), so we want the _____ to be small. There are two common ways to decrease the margin of error.

Factors that Affect Margin of Error

Assuming everything else remains the same, the margin of error will be smaller when…

1) The sample size is larger.

Margin of error = $t * \frac{s}{\sqrt{n}}$

For a population mean, the width of the interval is proportional to $\frac{1}{\sqrt{n}}$. This means that quadrupling the _____ size will cut the margin of error in half.

2) The confidence level is smaller.

Margin of error = $\boldsymbol{t} * \frac{s}{\sqrt{n}}$

The critical value t^* for 90% confidence is _____ (for all degrees of freedom) than the critical value t^* for 95% confidence, making the 90% CI narrower than the 95% CI.

What Should We Take Away?

How do we interpret the confidence **level** for a confidence interval for a population mean?

In _____ random sampling with the _____ sample _____, approximately ____ of ____ confidence intervals will _____ the population mean.

How do the sample size and confidence level affect the margin of error for a confident interval for a population mean?

Assuming everything else stays the same:

- **Increasing the _____ will decrease the _____.**
- **Increasing the _____ will _____ the margin of error.**

Name_____

AP Statistics CED 7.3 Daily Video 3 (Skill 4.B)

Justifying a Claim About a Population Mean Based on a Confidence Interval

What Will We Learn?

How do we construct and interpret confidence interval for a population mean?

2013 #1

An environmental group conducted a study to determine whether crows in a certain region were ingesting food containing unhealthy levels of lead. A biologist classified lead levels greater than 6.0 parts per million (ppm) as unhealthy. The lead levels of a random sample of 23 crows in the region were measured and recorded. The data are shown in the stemplot.

Lead Levels

2	8
3	0
3	5 8 8
4	1 1 2
4	6 8 8
5	0 1 2 2 3 4
5	9 9
6	3 4
6	6 8

Key: 2|8 = 2.8 ppm

(a) What proportion of crows in the sample had lead levels that are classified by the biologist as unhealthy?

(b) The mean lead level of the 23 crows in the sample was 4.90 ppm and the standard deviation was 1.12 ppm. Construct and interpret a 95 percent confidence interval for the mean lead level of crows in the region.

2013 #1

(a) What proportion of crows in the sample had lead levels that are classified by the biologist as unhealthy?

Circle the value that are greater than 6.0 ppm.

The proportion of crows in the sample that were classified as unhealthy is _____ ≈ _____.

Lead Levels

2	8
3	0
3	5 8 8
4	1 1 2
4	6 8 8
5	0 1 2 2 3 4
5	9 9
6	3 4
6	6 8

Key: 2|8 = 2.8 ppm

2013 #1

(b) The mean lead level of the 23 crows in the sample was 4.90 ppm and the standard deviation was 1.12 ppm. Construct and interpret a 95 percent confidence interval for the mean lead level of crows in the region.

State: Create a _____ confidence interval for ____ = mean lead of crows in the region.

Identify Procedure: _____.

Check Conditions:

Random: _____.

10% Condition: _____.

Approximately Normal: _____

_____.

The conditions are _____.

Lead Levels

2	8
3	0
3	5 8 8
4	1 1 2
4	6 8 8
5	0 1 2 2 3 4
5	9 9
6	3 4
6	6 8

Key: 2|8 = 2.8 ppm

Name_____

2013 #1

(b) The mean lead level of the 23 crows in the sample was 4.90 ppm and the standard deviation was 1.12 ppm. Construct and interpret a 95 percent confidence interval for the mean lead level of crows in the region

Calculate the Interval using CI = $\bar{x} \pm t * \frac{s}{\sqrt{n}}$ and then plug in appropriate values:

CI = _____ ± _____(_____)

Remember to use technology to find the critical value of *t or (t*)*.

= _____ ± _____

= _____ to _____

2013 #1 Interpret the Interval

"We can be _____ confident that the _____ from _____ to _____ ppm capture the _____ lead level in the population of _____ crows in this region.

What Should We Take Away?

How do we construct and interpret confidence interval for a population mean?

Make sure to:

- Define the _____ you are trying to _____.
- Identify the _____ you are using.
- Verify that the _____ for the procedure are met (with evidence!).
- Calculate the _____.
- Interpret the interval _____.

(You do _____ need to interpret the confidence _____ unless specifically asked.)

Name_____

AP Statistics CED 7.4 Daily Video 1 (Skill 1.F)

Setting up a Test for a Population Mean

What Will We Learn?

How do we state a null hypothesis for a test for a population mean?

How do we state an alternative hypothesis in a test for a population mean?

Some Reminders….

There are two major types of statistical inference in Units 6 – 9:

Confidence Intervals

- used to _____ the value of a _____.
- an _____ - rather than a single _____ - used to estimate a _____ to account for _____ variability

Significance Tests

- used to _____ about the value of a population parameter
- assess whether the _____ supporting a claim is likely or unlikely to happen by _____.

The Structure of Unit 7

In Unit 7, we will focus on _____ data that can be summarized by calculating the _____ in a sample or treatment group.

Topics 7.2 – 7.3 ⟶ _____ for a population _____

Topics 7.4 – 7.5 ⟶ _____ for a population _____

Topics 7.6 – 7.7 ⟶ _____ for a difference in _____

Topics 7.8 – 7.9 ⟶ _____ for a difference in _____

Got Hops?

An article on the internet claims that the average vertical jump for teens is 15 inches. Intrepid AP Statistics students at one large high school selected a random sample of 20 students and measured their vertical jumps. The students in the sample jumped an average of 15.8 inches, with a standard deviation of 2.33 inches. Do these data provide convincing evidence that the average vertical jump for all students at this school differs from 15 inches?

11.5	12.5	13.25	13.75	13.75	14.25	14.5	15.25	15.25	15.5
15.75	16.25	16.5	16.75	17.0	17.5	18.5	18.75	19.0	20.5

Null Hypothesis

In a statistical test, the _____ hypothesis is often a claim of "_____" or "_____".

In the vertical jump example, the null hypothesis is that the _____ vertical jump for students at _____ high school is _____ different than the average for _____ teens.

H_0: μ = _____, where μ = the _____ vertical jump for _____ students at this high school.

Until we have _____ evidence otherwise, we assume H_0 is correct.

Name_____

Alternative Hypothesis

In a statistical test, the _____ hypothesis is the claim that we _____ to support with _____ from the data collected.

In the vertical jump example, the AP Statistics students want to know if the average jump for students at this school _____. So, the alternative hypothesis is the mean vertical jump for all students at this high school is _____.

H_0: μ ____ 15

H_a: μ ____ 15 where μ = the mean vertical jump for _____

Stating Hypothesis: Summary

For hypotheses about a population mean:

- The _____ is a statement of equality, typically H_0: μ = [_____].
- The _____ always contains a _____ inequality, typically
  * when the inequality is ____ or ____, the alternative is called "_____."
  * when the inequality is ____, the alternative is called "_____."
  * The choice of alternative is determined by the _____ of interest and should be stated _____ data collection begins.
- Never refer to _____ (such as ____) in the hypotheses!
- Remember to define _____.

I Can Go for Miles and Miles

A tire manufacturer must test its Tread40 tires to see if they last **more than 40,000 miles,** on average. A quality control engineer selects a random sample of 35 tires and puts them on a machine that simulates driving conditions until the tread on the tire is 2/32" when they must be replaced. The average mileage for the sample is 42,348 miles with a standard deviation of 2,140 miles.

State appropriate hypotheses for the engineer's test.

H_0: _____

H_a: _____ where μ = _____

What Should We Take Away?

How do we state a null hypothesis for a test for a population mean?

H_0: μ ____ [null value], where μ = the _____ mean defined in _____.

How do we state an alternative hypothesis in a test for a population mean?

H_a: μ _____ [null value]
H_a: μ _____ [null value]
H_a: μ _____ [null value]

Remember that the alternative hypothesis will take one of three forms and should be determined by the research question. It should always be stated before the research begins.

Name_____

AP Statistics CED 7.4 Daily Video 2 (Skill 4.C)

Setting Up a Test for a Population Mean

What Will We Learn?

How do we identify an appropriate significance test procedure for a population mean?
How do we verify the conditions for performing a significance test for a population mean?

Got Hops?

An article on the internet claims that the average vertical jump for teens is 15 inches. Intrepid AP Statistics students at one large high school selected a random sample of 20 students and measured their vertical jumps. The students in the sample jumped an average of 15.8 inches, with a standard deviation of 2.33 inches. Do these data provide convincing evidence that the average vertical jump for all students at this school differs from 15 inches?

11.5	12.5	13.25	13.75	13.75	14.25	14.5	15.25	15.25	15.5
15.75	16.25	16.5	16.75	17.0	17.5	18.5	18.75	19.0	20.5

Got Hops?

In a previous video, we stated the hypotheses:

H_0: μ ____ 15

H_a: μ ____ 15 where μ = the mean vertical jump for _____

Identifying the Procedure

When the goal is to test a claim about a population mean, we use a _____ _____

Checking the Conditions

Remember that for _____ inference procedures in AP Statistics you must verify that the _____ for using that procedure are _____.

In general, you should always check for:

- _____ in the methods used to collect data, and
- that the appropriate _____ distribution has the correct _____.

Checking the Conditions

Here are the conditions for a _____ for a population mean.

To check for independence:

1. The data are collected using a _____ sample from the population or a _____ experiment.
2. When sampling _____ replacement, the _____ is less than or equal to _____ of the population size.

To check that the shape of the _____ distribution is approximately normal:

3. The samples size is _____ ($n \geq$ ____) OR if $n < 30$, the sample data are free from strong _____ and _____.

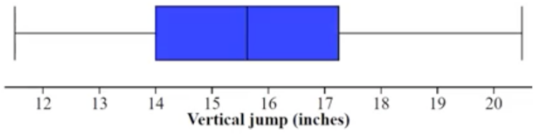
Name_____

Checking the Conditions

Does the average vertical jump of all students at a large high school differ from 15 inches? Here are the vertical jumps of 20 randomly selected students. Check if the conditions for calculating a confidence interval are met.

11.5	12.5	13.25	13.75	13.75	14.25	14.5	15.25	15.25	15.5
15.75	16.25	16.5	16.75	17.0	17.5	18.5	18.75	19.0	20.5

1. The _____ students were _____ selected.
2. It's reasonable to believe _____ students is less than or equal to _____ of all students at a large high school.
3. A boxplot of the vertical jumps shows no strong _____ or _____.



The conditions are _____.

(*Note you could make a dot plot, histogram or boxplot, but you must provide evidence!!)

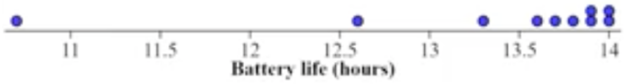
Long Live the Tablet!

CB Tablets claims that its tablet computers have an average battery life of 14 hours under normal usage. A consumer advocacy group wonders if the batteries have shorter lives, on average, than the claim by CB Tablets. The table shows the battery life (in hours) for 10 tablets.

10.7	12.6	13.3	13.6	13.7
13.8	13.9	13.9	14.0	14.0

Check if the _____ for performing a significance test are met. (Use ✓ 's)

1. The 10 tablets are _____ selected.
2. It's reasonable to believe _____ tablets is less than or equal to _____ of all tablets made by CB Tablets.
3. A dotplot of the battery life data shows strong _____ and a potential _____.



The conditions are _____

What Should We Take Away?

How do we identify an appropriate significance test procedure for a population mean?

A _____

How do we verify the conditions for performing a significance test for a population mean?

1. The data are collected using a _____ sample from the population or _____ experiment.
2. When sampling _____ replacement, the sample _____ is less than or equal to _____ of the population size.
3. The sample size is _____ ($n \geq$ ____) OR If $n < 30$, the sample data are _____ from strong _____ or _____.

Name_____

AP Statistics CED 7.5 Daily Video 1 (Skill 3.E)

Carrying Out a Test for a Population Mean

What Will We Learn?

How do we calculate an appropriate test statistic in a test for a population mean?
How do we calculate a *p*-value in a test for a population mean?

Got Hops?

An article on the internet claims that the average vertical jump for teens is 15 inches. Intrepid AP Statistics students at one large high school selected a random sample of 20 students and measured their vertical jumps. The students in the sample jumped an average of 15.8 inches, with a standard deviation of 2.33 inches. Do these data provide convincing evidence that the average vertical jump for all students at this school differs from 15 inches?

11.5	12.5	13.25	13.75	13.75	14.25	14.5	15.25	15.25	15.5
15.75	16.25	16.5	16.75	17.0	17.5	18.5	18.75	19.0	20.5

Got Hops?

From previous videos: H_0: μ ____ 15 and H_a: μ ____ 15
where μ = the mean vertical jump for all students at this school. Conditions are _____.

Calculating a Test Statistic

In the vertical jump study, $\bar{x}$ = _____
This is evidence for H_a: _____ because $\bar{x}$ = 15.8 ____ 15. We want to know how likely it is to get evidence for H_a the _____ or _____ by chance along when H_0 is _____. After verifying that the _____ are met, calculate the standardized test statistic:

standardized test statistic = $\frac{\text{_____} - \text{_____}}{\text{_____}}$

Calculating a Test Statistic

For a _____ for a population mean, the standardized test statistic is:

$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ where the *t* statistic has n approximate *t* distribution with df = *n* -1

From technology: _____ inches and _____ inches, _____

Plug in the values and find $t = \frac{\text{–}}{\sqrt{\text{___}}}$ = _____ with df = _____ = ____

Calculating a Test Statistic

The components for the test statistic formula can be found on the formula sheet. Locate them now.

Calculating a *p*-value

The *p*-value is the _____ of observing a test statistic as _____ or more _____ than the observed test statistic when the _____ hypothesis and probability model are _____ to be _____.

For a _____ for a population mean, the *p*-value is calculated from a _____ distribution with _____ degrees of freedom.

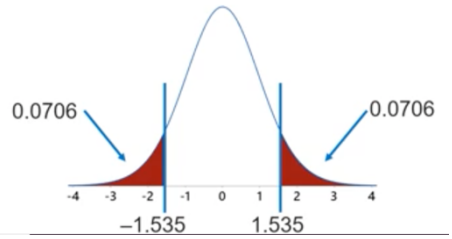
Name_____

Calculating a *p*-value

Remember: _____ with df = _____

Because our alternative hypothesis is _____, we want to find:

$P(t \leq$ _____$) + P(t \geq$ _____$) \longrightarrow 2 \cdot P(t \geq |$_____$|)$



Using technology, *p*-value = _____ = _____

STOP: Follow the steps for the TI-84 or use Table B. (Technology will be more accurate!)

I Can Go for Miles and Miles

A tire manufacturer must test its Tread40 tires to see if they last **more than 40,000 miles,** on average. A quality control engineer selects a random sample of 35 tires and puts them on a machine that simulates driving conditions until the tread on the tire is 2/32" when they must be replaced. The average mileage for the sample is 42,348 miles with a standard deviation of 2,140 miles. Calculate the standardized test statistic and the *p*-value.

I Can Go for Miles and Miles

From previous videos we know:

H_0: _____

H_a: _____ where _____ = mean milage for _____ Tread40 tires

Conditions are _____.

Calculate a Test Statistic

$\bar{x}$ = _____, s = _____ and n = _____

plug in values to find the test statistic:

$$t = \frac{\_\_\_\_ - \_\_\_\_}{\frac{\_\_\_\_}{\sqrt{\_\_}}} = \_\_\_\_\_ \text{ with df} = \_\_\_\_\_ = \_\_\_\_$$

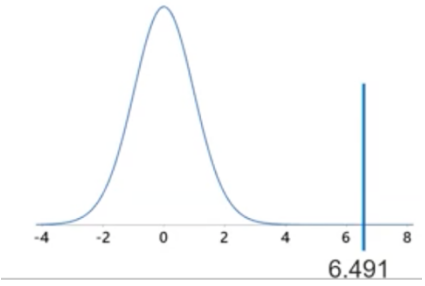
Calculating the *p*-value

t = _____ with df = _____;

Because our alternative hypothesis is: H_a: _____,

we want to find $P(t \geq$ _____$)$

Using technology or Table B the *p*-value is _____



What Should We Take Away?

How do we calculate an appropriate test statistic in a test for a population mean?

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

How do we calculate a *p*-value in a test for a population mean?

- If H_a: _____ [null value], _____ = $P(t$ ____ observed test statistic)
- If H_a: _____ [null value], _____ = $P(t$ ____ observed test statistic)
- If H_a: _____ [null value], _____ = $2 \cdot P(t$ ____ observed test statistic)

Name_____

AP Statistics CED 7.5 Daily Video 2 (Skill 4.B)

Carrying Out a Test for a Population Mean

What Will We Learn?

How do we interpret the *p*-value in a test for a population mean?

How do we state a conclusion for a significance test for a population mean?

Got Hops?

An article on the internet claims that the average vertical jump for teens is 15 inches. Intrepid AP Statistics students at one large high school selected a random sample of 20 students and measured their vertical jumps. The students in the sample jumped an average of 15.8 inches, with a standard deviation of 2.33 inches. Do these data provide convincing evidence that the average vertical jump for all students at this school differs from 15 inches?

11.5	12.5	13.25	13.75	13.75	14.25	14.5	15.25	15.25	15.5
15.75	16.25	16.5	16.75	17.0	17.5	18.5	18.75	19.0	20.5

Got Hops?

From previous videos: H_0: μ ____ 15 and H_a: μ ____ 15

where μ = the mean vertical jump for all students at this school. Conditions are met.

$\bar{x}$ = _____, s = _____, t = _____with df = _____, p-value = _____

The p-value measures how likely it is to get evidence for H_a as _____ or _____ than the observed evidence by _____ when H_0 is _____.

Interpreting a *p*-value

Assuming _____, there is _____ of getting a sample mean of _____ or _____, by _____ alone in the _____ sample (_____ assignment).

Assuming the _____ is 15 inches, there is a _____ of getting a _____ as extreme as or _____ than 15.8 inches in either direction, by _____ alone in a _____ sample of _____ students.

Stating a Conclusion

Small *p*-values ⟶ test statistic is _____ to occur by random chance alone.

Large *p*-values ⟶ test statistic is _____ to occur by random chance alone.

- Because the *p*-value of _____ ≤ α = _____, we reject H_0.
  There is _____ that [state H_a _____]
- Because the *p*-value of _____ ≤ α = _____, we fail to reject H_0.
  There is _____ that [state H_a _____]

Name_____

Stating a Conclusion

No significance level was stated in the vertical jump example, so we'll use α = _____, which is the most common significance level.

Because the *p*-value of _____ > α = _____, we _____ H_0.
There is _____ statistical evidence that the mean vertical jump for _____ is different than 15 inches.

I Can Go for Miles and Miles

A tire manufacturer must test its Tread40 tires to see if they last **more than 40,000 miles,** on average. A quality control engineer selects a random sample of 35 tires and puts them on a machine that simulates driving conditions until the tread on the tire is 2/32" when they must be replaced. The average mileage for the sample is 42,348 miles with a standard deviation of 2,140 miles. Calculate the standardized test statistic and the *p*-value.
The *p*-value for a one-sample *t* test of the mean milage is 9.99×10^{-8}.
Interpret the *p*-value and make a conclusion at the $\alpha = 0.01$ significance level.

Tire Mileage: Interpret *p*-value

H_0: μ ____ 40,000 miles
H_a: μ ____ 40,000 miles, where μ = mean milage for all Tread40 tires.

$\bar{x}$ = _____, *s* = _____, *t* = _____with df = _____, *p*-value = _____
Interpret the *p*-value.
Assuming the mean _____, there is a _____ of getting a _____ mean mileage of 42,348 miles or _____, by _____ alone in a _____ sample of 35 tires.

Tire Milage: Conclusion

Make a conclusion at the $\alpha = 0.01$ significance level.
Because the _____ of $9.99 \times 10^{-8} \leq \alpha = 0.01$, we _____ H_0.
There is convincing _____ that the mean mileage for _____ _____ is _____ than 40,000 miles.

What Should We Take Away?

How do we interpret the *p*-value in a test for a population mean?
Assuming _____, there is _____ of getting a sample mean of <*observed difference*> or <*greater/less/more different*> by _____ alone in the _____ sample (random assignment).
How do we state a conclusion for a significance test for a population mean?

- **Because the *p*-value of _____ $\leq \alpha$ = _____, we reject H_0.**
  There is _____ that [state H_a _____]
- **Because the *p*-value of _____ $\leq \alpha$ = _____, we fail to reject H_0.**
  There is _____ that [state H_a _____]

Name_____

AP Statistics CED 7.5 Daily Video 3 (Skill 4.E)

Carrying Out a Test for a Population Mean

What Will We Learn?

How do we perform a complete significance test for a population mean?

Bakin' Bacon!

Doug likes his bacon crispy, but still tender. Bridget recommends that Doug try adding mango habanero seasoning to his bacon during frying because the sugar in the seasoning will caramelize, creating a crust that helps the bacon retain moisture.

To test this claim, Doug randomly assigned half of one package of bacon to be cooked without seasoning. The remaining half-package was seasoned with a mango habanero seasoning and cooked for the same length of time. After cooking, Doug measured the weight (in grams) of each half-package of cooked bacon to see if the seasoning helped the bacon retain its moisture. Then he repeated this process for 9 more packages. (Yum!)

Paired Data

Here are the data.

Because these data were produced from a _____ experiment, they are an example of _____.

Package	Weight With Seasoning (grams)	Weight Without Seasoning (grams)
1	278	260
2	268	238
3	271	281
4	301	306
5	256	240
6	248	244
7	240	220
8	260	252
9	286	287
10	309	294

Paired Data

Matched pairs can be thought of as a _____ sample of pairs.

To analyze _____ data, find the _____ in each pair.

Package	Weight With Seasoning (grams)	Weight Without Seasoning (grams)	Difference in Weight (with – without)
1	278	260	18
2	268	238	30
3	271	281	−10
4	301	306	−5
5	256	240	16
6	248	244	4
7	240	220	20
8	260	252	8
9	286	287	−1
10	309	294	15

Paired Data

Once differences between _____ of values are found, compute the _____.

From technology:

$\bar{x}_d$ = _____

s_d = _____

n = _____

Package	Weight With Seasoning (grams)	Weight Without Seasoning (grams)	Difference in Weight (with – without)
1	278	260	18
2	268	238	30
3	271	281	−10
4	301	306	−5
5	256	240	16
6	248	244	4
7	240	220	20
8	260	252	8
9	286	287	−1
10	309	294	15

Inference (confidence intervals and significance tests) can be performed about μ_d just as for a population mean.

Note: Make sure you can calculate these statistics on technology.

Bakin' Bacon!

Doug randomly assigned half of a package of bacon to be cooked without seasoning. The remaining half was seasoned with a mango habanero seasoning and cooked for the same length of time. The

Name_____

table shows the difference in weigh (with seasoning – without seasoning) of each half-package of cooked bacon for 10 packages. Do these data give convincing evidence that the seasoning causes cooked bacon to **retain more weight**, on average, for packages of bacon like the ones in this study?

Difference in Weight (with – without)
18
30
–10
–5
16
4
20
8
–1
15

H_0: _____

H_a: _____

where μ_d = the _____ in weight (with seasoning – without seasoning) for packages of cooked bacon like the ones in this study

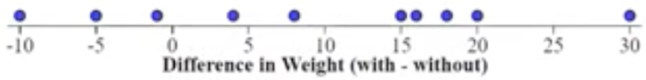
No significance level is stated, so we'll use $\propto$ = _____.

Bakin' Bacon

We will perform a _____

Conditions

- Half-packages were _____ assigned to be cooked with or without seasoning
- No 10% condition! (Because we are ____ randomly sampling _____ replacement.)
- The dotplot of differences shows ____ strong skewness or outliers.



Find *t* statistic:

From technology:

$\bar{x}_d$ = _____

s_d = _____

n = _____

Plug in the values to determine the *t* statistic:

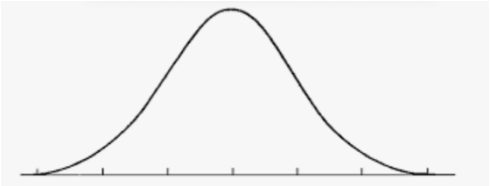
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Calculate the *p*-value

Using technology or Table B:

p-value = *P*(*t* _____) = _____

Label the graph below



Interpret the p-value

Because the *p*—value of _____, we _____ H_0.

There is _____ statistical _____ that the seasoning causes cooked bacon to retain more weight, on average for packages of bacon _____ in this study.

What Should We Take Away?

How do we perform a complete significance test for a population mean?

Make sure to:

- State the null and alternative hypotheses and _____.
  (For _____, indicate the direction of the _____.)
- Identify the _____.
- Identify the _____ you are using.
- Verify that the conditions for the procedure are met (_____).
- Calculate the _____ and _____.
- Make a conclusion based on the _____.
  (You do not need to interpret the p-value specifically unless specifically asked)

Name_____

AP Statistics CED 7.6 Daily Video 1 (Skill 4.C)

Confidence Intervals for the Difference of Two Means

What Will We Learn?

How do we identify an appropriate confidence interval procedure for the difference of two population means?

How do we verify the conditions for calculating a confidence interval for the difference of two population means?

Arachnophobia?

Mammal species typically have larger males than females. However, there are some species in the animal kingdom where the females are larger than the males. This most notably occurs with spiders. To determine the difference in the mean body lengths of female and male spiders researchers collected a random sample of 14 adult female spiders and a random sample of 14 adult male spiders from the genus *Argiope*, a particularly large type of spider.

Arachnophobia

The data on body lengths are shown below. Calculate and interpret a 95% confidence interval for the difference in the population mean body lengths of female and male *Argiope* spiders.

Female spider body lengths (in millimeters)

16.3	15.3	15.3	17.4	19.7	10.3	12.6
9.8	11.4	7.3	15.4	17.0	12.7	17.6

Male spider body lengths (in millimeters)

4.1	5.6	5.5	3.6	4.6	4.3	3.5
4.5	2.9	2.8	3.2	4.0	3.6	5.3

Identifying the Procedure

How many groups? We are comparing _____ of spiders (female and male).
What type of data were collected? Measurements of body lengths in millimeters (_____).
What are we asked to do? Estimate the difference in the population _____ body lengths of female and male spiders with a 95% <u>confidence interval.</u>

2-sample *t*-interval for the difference in population means

Checking the Conditions

To check for independence:

1. The data should come from _____ independent _____ samples OR a _____ experiment.
2. When sampling _____ replacement, the samples should be _____ or _____ 10% of their respective populations.

To check that the shape of the sampling distribution of $(\bar{x}_1 - \bar{x}_2)$ is approximately normal:

3. Both n_1 and n_2 should be greater than _____ OR both _____ should come from approximately _____ distributed populations.
   - If samples are small, the shapes of the sample distributions should be _____ from extreme _____ or _____.

Checking the Conditions (Remember to ✓ your conditions)

To determine the difference in the mean body lengths of female and male spiders, researcher collected a random sample of 14 adult female spiders and a random sample of 14 adult male spiders from the genus *Argiope*.

1. Both samples of spiders were _____ selected.
2. 14 female and 14 male spiders are certainly _____ 10% of the population of these spiders.
3. Both samples have _____ observations so we must verify there is no extreme _____ or _____. (Sketch the histograms next to each table.)

Name_____

Female spider body lengths (in millimeters)

16.3	15.3	15.3	17.4	19.7	10.3	12.6
9.8	11.4	7.3	15.4	17.0	12.7	17.6

Male spider body lengths (in millimeters)

4.1	5.6	5.5	3.6	4.6	4.3	3.5
4.5	2.9	2.8	3.2	4.0	3.6	5.3

4. Both samples are roughly _____ and _____, so it is reasonable to assume they come from _____ populations.

Checking the Conditions (Remember to ✓ your conditions)

To determine the difference in the mean body length of female and male spiders, researchers actually collected a random sample of 35 adult female spiders and a random sample of 35 adult male spiders from the genus *Argiope*. If the researchers wanted to calculate a 95% confidence interval for the difference (*female – male)* in mean body length of *Argiope* spiders, are the conditions met?

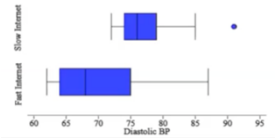
1. Both samples of spiders were _____ selected.
2. _____ female and _____ male spiders are certainly _____ to 10% of the populations of _____.
3. Both samples have more the ____ observations so the _____ distribution of ($\bar{x}_1$ - $\bar{x}_2$) is approximately _____.

Checking the Conditions: Multiple Choice

A teacher wanted to measure the effects of slow internet on student stress levels. She randomly assigned 18 students from her class to complete an online task in a room where the internet connection was slow due to the distance from the Wi-Fi hub. The remaining 18 students completed the same online task in a room very close to the hub which allowed much faster connection. The diastolic blood pressure for each student was measured immediately at the conclusion of the task. The teacher would like to calculate a 95% confidence interval for the difference in mean blood pressure of students with fast and slow internet connections. Summary statistics and bloxplots of the blood pressure readings are given.

Summary Statistics

Group Name	n	mean	SD	min	Q_1	med	Q_3	max
1: Slow Internet	18	77.111	4.676	72	74	76	79	91
2: Fast Internet	18	69.722	6.935	62	64	68	75	87



for a

Which of the following statements is true regarding the teacher's method?

(A) The students were not randomly selected so it is not appropriate to calculate a confidence interval for a difference of means.

(B) 36 students is a large enough sample for the sampling distribution of the difference in sample means to be approximately normal.

(C) There is clear skewness and an outlier in the sample distributions so the teacher cannot assume the data come from approximately normally distributed populations.

(D) There is clear skewness and an outlier in the sample distributions so the two groups cannot be assumed to be independent.

(E) The conditions appear to be verified so the teacher can proceed with her chosen inference method.

What Should We Take Away?

How do we identify an appropriate confidence interval procedure for the difference of two population means?

When estimating the _____ in means from two populations, use a _____ _____ for the difference in _____ means.

How do we verify the conditions for calculating a confidence interval for the difference of two population means?

1. **The data come from two _____ samples or a _____ experiment.**
2. **When sampling _____ replacement, the samples should be _____ than or _____ 10% of their respective _____.**
3. **Both n_1 and n_2 should be _____ OR both _____ should come from _____.**

Name_____

AP Statistics CED 7.6 Daily Video 2 (Skill 4.E)

Confidence Intervals for the Difference of Two Means

What Will We Learn?

How do we determine the margin of error when estimating a difference in means?

How do we calculate a confidence interval for a difference of means?

Formula for Success!

2-sample *t*-interval for the difference in population means

Confidence interval: statistic ± (_____)(_____)

(*Stop and locate the formula on the AP Statistics Formula Sheet*)

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

CI = _____ ± _____

The margin of error describes how much a value of a _____ is likely to vary from the value of the corresponding _____.

Calculating the Margin of Error

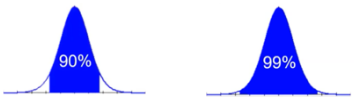
CI = point estimate ± _____

(_____)(_____)

The critical value is determined by how _____ we want our estimate to be.

$$t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The standard error of the statistic tells how much the statistics typically _____ from the _____



Eight Legged Friends

Mammal species typically have larger males than females. However, there are some species in the animal kingdom where the females are larger than the males. This most notably occurs with spiders. To determine the difference in the mean body lengths of female and male spiders researchers collected a random sample of 14 adult female spiders and a random sample of 14 adult male spiders from the genus *Argiope*, a particularly large type of spider.

Female spider body lengths (in millimeters)

16.3	15.3	15.3	17.4	19.7	10.3	12.6
9.8	11.4	7.3	15.4	17.0	12.7	17.6

Male spider body lengths (in millimeters)

4.1	5.6	5.5	3.6	4.6	4.3	3.5
4.5	2.9	2.8	3.2	4.0	3.6	5.3

Calculating the Confidence Interval

Using the calculator, find the sample means and standard deviations and then plug into the formula:

$$(\bar{x}_F - \bar{x}_M) \pm t^* \sqrt{\frac{s_F^2}{n_F} + \frac{s_M^2}{n_M}}$$

$\bar{x}_F$ = _____ $\bar{x}_M$ = _____

s_F = _____ s_M = _____

What About the Degrees of Freedom?

In Topic 7.2 you learned about finding the critical value from a _____ using df = n – 1 for a 1-sample *t*-interval. In a _____ the degrees of freedom are found using technology. Follow the calculator steps on the next slides to find your degrees of freedom, 2-sample *t* interval and other information. Write your confidence interval with df in the box above.

Name_____

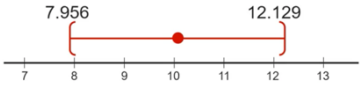
Finding the Critical Value

Sometimes you will have to find the critical value for a confidence interval. The critical value can be found using technology. Follow and perform the steps on your calculator as you watch the video.

For our example, the critical value t = invT (0.25, 14.73) = _____

*NOTE: Remember to only use an area equal to the area of **ONE** tail.

Summing up the CI Calculations



Point Estimate: _____ = _____ = _____

Margin of Error: _____ = _____ OR _____ $\sqrt{\quad + \quad}$

95% Confidence Interval: _____

Summing Up the CI Calculations

You will need to be familiar with all forms of how a confidence interval can be written:

(_____, _____) OR _____ ± _____ OR (_____)±_____ $\sqrt{\quad + \quad}$

Multiple-Choice Example

Three high school students wanted to know what type of stimulus, visual or auditory, resulted in faster reaction time in young drivers. In a group of 25 volunteers who were teenage drivers, the students randomly assigned 12 to complete a visual reaction time test and the remaining 13 were assigned to complete an auditory reaction time test. The summary statistics are shown in the table to the right.

	n	Mean (in milliseconds)	Standard deviation (in milliseconds)
Visual Reaction Time Test	12	256	35
Auditory Reaction Time Test	13	224	23

(STOP the video and calculate the CI!)

HINT: CI = statistic ± margin of error (Look back in your notes at the example!)

Which of the following gives the correct calculation for a 90% confidence interval for the difference in mean reaction times for visual and auditory reaction time tests in teenage drivers?

(A) $(256-224) \pm 1.645\sqrt{\frac{35^2}{12}+\frac{23^2}{13}}$

(B) $(256-224) \pm 1.711\sqrt{\frac{35^2}{12}+\frac{23^2}{13}}$

(C) $(256-224) \pm 1.73\sqrt{\frac{35^2}{12}+\frac{23^2}{13}}$

(D) $(256-224) \pm 1.73\sqrt{\frac{35}{12}+\frac{23}{13}}$

(E) $(256-224) \pm 1.96\sqrt{\frac{35^2}{13}+\frac{23^2}{12}}$

Cross out the options as you watch the video. Then, follow the steps on your calculator to find the critical value for t to select the correct answer.

What Should We Take Away?

How do we determine the margin of error when estimating a difference in means?

Margin of error = (_____)(_____) where margin of error = $t^*\sqrt{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}$

How do we calculate a confidence interval for a difference of means?

CI = Point Estimate ± Margin of Error OR CI = (_____)±_____ $\sqrt{\quad + \quad}$

***Note: degrees of freedom are found using technology!!**

Name_____

AP Statistics CED 7.7 Daily Video 1 (Skill 4.B)

Justifying a Claim with a Confidence Interval for Difference of Means

What Will We Learn?

How do we interpret a confidence interval for a difference in means?

How do we justify a claim based on a confidence interval for a difference in means?

Spiders, The Final Episode

Mammal species typically have larger males than females. However, there are some species in the animal kingdom where the females are larger than the males. This most notably occurs with spiders. To determine the difference in the mean body lengths of female and male spiders researchers collected a random sample of 14 adult female spiders and a random sample of 14 adult male spiders from the genus *Argiope*, a particularly large type of spider. The data on body lengths are shown to the right. Calculate and iterpret a 95% confidence interval for the difference in the population mean body lengths of female and male *Argiope* spiders.

Female spider body lengths (in millimeters)

16.3	15.3	15.3	17.4	19.7	10.3	12.6
9.8	11.4	7.3	15.4	17.0	12.7	17.6

Male spider body lengths (in millimeters)

4.1	5.6	5.5	3.6	4.6	4.3	3.5
4.5	2.9	2.8	3.2	4.0	3.6	5.3

Interpreting the Confidence Interval

In general, here is how to interpret a confidence interval for a _____ parameter:

"We are *C*% confident that the interval from _____ to _____ captures the [value to be estimated]."

From Topic 7.6, Video 2, the 95% confidence interval for the *Argiope* spiders example is _____ to _____. So, we would say:

"We are _____ that the interval from _____ to _____ captures the difference (_____) in the true mean body lengths of _____ _____."

Alternate Interpretation:

" I am _____ that the true mean body length of _____ is between 7.956 and 12.129 mm _____ the mean body length of _____ _____."

If we estimate the _____ in the mean body lengths by _____ in the opposite order (_____), the 95% confidence interval would be (_____).

Notice the change in the interpretation:

"I am _____ that the true mean body length of _____ is between 7.956 and 12.129 mm _____ the mean body length of _____ _____."

Incorrect Interpretations

"We are 95% confident that the interval from 7.956 to 12.129 mm captures the difference (*female – male*) in the mean body lengths of _____ *Argiope* spiders.

NOTE: "these" refers to the _____ instead of the _____.

"There is a 95% _____ that the interval from 7.956 to 12.129 mm captures the difference (*female – male*) in the true mean body lengths of female and male *Argiope* spiders."

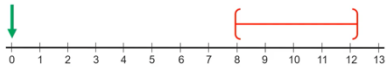
The word "chance" implies probability and the confidence level is not a probability of capturing the population parameter. Any confidence interval either includes the parameter or does not include the parameter.

Name_____

Justifying a Claim

The researchers in this study hypothesize that male spiders tend to be smaller than female spiders so that they can maneuver around webs and habitats faster against the effects of gravity. This makes finding a mate easier. To further study this phenomenon, they wish to identify various spider species where the females are larger than the males, on average. Is there convincing evidence that *Argiope* spiders should be included in this follow up study?

Justifying a Claim



If female and male *Argiope* spiders were _____ in size, the difference in their mean body lengths would be _____. Because _____ is not in the 95% confidence interval (_____), 0 is not a _____ value for $\mu_F - \mu_M$, the difference in _____ body lengths of female and male *Argiope* spiders. Therefore, we _____ have convincing _____ that female *Argiope* spiders have _____ larger body lengths than males, _____. *Argiope* spiders _____ be included in the follow-up study.

Multiple Choice Example

A restaurant manager was trying to decide whether to switch the current take-out food containers from foam to plastic. The plastic containers would be a more expensive option, but the supplier claims that plastic containers retain heat better than foam containers, on average. The manager cooked 16 identical dishes and randomly assigned 8 or them to foam containers and the other 8 to plastic containers. After one hour, the manager measured the internal temperature of each meal (in degrees Fahrenheit) to see which type of container more effectively maintain the food's heat. A 95% confidence interval for the difference in the mean food temperatures (*foam – plastic*) was calculated to be (-9.3, 3.2). Which of the following conclusions can the restaurant manager make?

A 95% confidence interval for the difference in mean food temperatures (*foam – plastic*) was calculated to be (-9.3, 3.2). Which of the following conclusions can the restaurant manager make?

A. The manager should be 95% confident that the mean food temperature in the 8 foam containers is between 3.2°F and 9.3°F higher than the mean food temperature in the 8 plastic containers.

B. The manager should be 95% confident that the interval from -9.3°F to 3.2°F captures the true difference in the mean internal temperatures of food in foam and in plastic containers.

C. The manager should switch to plastic containers because 95% of them helped the food maintain heat better than in the foam containers.

D. Since 0 is not on the interval, there is convincing evidence that plastic helps food maintain heat better than in foam containers, on average. The manager should switch to plastic containers.

E. Since 0 is on the interval, there is convincing evidence that foam helps food maintain heat better than in plastic containers, on average. The manager should switch to plastic containers.

What Should We Take Away?

How do we interpret a confidence interval for a difference in means?

"We are *C*% confident that the interval from _____ to _____ captures the [value to be estimated]."

How do we justify a claim based on a confidence interval for a difference in means?

- **If _____ the values in the confidence interval are _____ with the claim, there is convincing evidence for the claim.**
- **If _____ of the values in the confidence interval are _____ with the claim, there is not convincing evidence for the claim.**

Name_____

AP Statistics CED 7.7 Daily Video 2 (Skill 4.D)

Justifying a Claim with a Confidence Interval for Difference of Means

What Will We Learn?

How do we construct and interpret a confidence interval for a difference of means?

AP Exam Pointers

- Read the _____ question, including _____ parts, for understanding.
- Then _____ the entire question for _____ and _____ important parts.
- Make sure to answer the _____ that is _____. If a _____ part of the question asks you to answer based on a _____ part, make sure to address it in the _____ part.
- Communicate _____! Define any _____ you use. Don't forget _____!

2009 Question 4

One of the two fire stations in a certain town responds to calls in the northern half of the town, and the other fire station responds to calls in the southern half of the town. One of the town council members believes that the two fire stations have different mean response time. Response time is measured by the difference between the time an emergency call comes into the fire station and the time the first fire truck arrives at the scene of the firs.

Data were collected to investigate whether the council member's belief is correct. A random sample of 50 calls selected from the northern fire station had a mean response time of 4.3 minutes with a standard deviation of 3.7 minutes. A random sample of 50 calls selected from the southern fire station had a mean response time of 5.3 minutes with a standard deviation of 3.2 minutes.

(a) Construct and interpret a 95 percent confidence interval for the difference in mean response times between the two fire stations.

(b) Does the confidence interval in part (a) support the council member's belief that the two fire stations have different mean response time? Explain. (Highlight key points as you watch video!)

2009 Question 4, Part (a)

(a) Construct and interpret a 95 percent confidence interval for the difference in mean response times between the two fire stations.

Define parameters: Let μ_N = _____

μ_S = _____

Identify procedures: _____

Check conditions: (Be sure to use a ✓ when check conditions!)

1. Randomness: _____

_____

2. ____ calls are likely less than or equal to _____ of _____ calls from the northern fire station.

____ calls are likely _____ 10% of all calls _____.

3. n_N = _____ and n_S = _____ so _____ samples are _____ that the _____ of $\bar{x}_N$ - $\bar{x}_S$ is _____ normal by the _____.

2009 Question 4, Part (a)

Mechanics: Using your calculator, run a 2-sample T interval as you watch the video. Then fill in the formula: (_____) $\pm$ _____ $\sqrt{\frac{\quad}{\quad}+\frac{\quad}{\quad}}$ The *t* critical value of *t*: _____

Confidence Interval = _____

Name_____

2009 Question 4, Part (a)

Interpretation:

I am _____ that the _____ in the population mean response times for the two fire stations (_____) is between -2.37 and 0.37 minutes.

2009 Question 4, Part (b)

(b) Does the confidence interval in part (a) support the council member's belief that the two fire stations have different mean response time? Explain.

From part (a), the 95% CI is (_____).

Since _____ is _____ the interval, it is a _____ value for the difference in the mean response time. Therefore, this confidence interval _____ support the council member's belief that there is a _____ for the two fire stations.

*Note: Do not say that the council member is _____!

2009 Question 4 Scoring

Section 1: Identify the appropriate confidence interval by _____ or _____ and check for appropriate _____.

Section 2: Show a correct confidence interval, either by displaying the numbers in the _____ or by writing the _____.

Section 3: Give a correct _____ of the confidence interval in _____.

Section 4: Make a correct conclusion _____, supported by the fact that _____ is contained within the _____ confidence interval.

2009 Question 4, What if….?

Sometimes you are asked to interpret the confidence LEVEL instead of the confidence interval. Suppose the AP Exam question had a part (c):

(c) Interpret the meaning of the 95% confidence level for this interval.

If _____ possible random samples of _____ from _____ fire station were selected and a 95% confidence interval was constructed from _____ of samples, then _____ of these intervals would succeed in _____ the difference (_____) in the mean response times for _____ from the two fire stations.

What Should We Take Away?

How do we construct and interpret a confidence interval for a difference of means?

Be sure to:

- ☐ **Define the difference in means you are trying to estimate: Indicate the _____ of the difference. Define any _____ you use.**
- ☐ **Identify the _____ you are using.**
- ☐ **Verify that the _____ for the procedure are _____.**
- ☐ **_____ the confidence interval.**
- ☐ **Interpret the interval _____.**

Name_____

AP Statistics CED 7.8 Daily Video 1 (Skill 1.F)

Setting Up a Test for The Difference of Two Population Means

What Will We Learn?

How do we state a null hypothesis in a test for a difference in means?

How do we state an alternative hypothesis in a test for a difference means?

Contagious Yawning

Three students wanted to test out the urban legend that yawning is contagious. From a group of 27 volunteers, they randomly assigned 14 people to have a children's nighttime store read to them while the storyteller yawned occasionally. The remaining 13 volunteers were read the same children's nighttime story, but without any yawning by the storyteller. The subjects in both groups were observed and the number of times each person yawned was recorded, as shown in the table below.

Yawn	7	3	7	6	8	6	5	6	3	5	6	7	4	5
No Yawn	3	2	2	2	0	1	2	1	1	4	3	0	5	

Is there convincing statistical evidence that people yawn more, on average, when watching someone yawn? Use an $\alpha = 0.05$ level of significance.

Null Hypothesis

If you were a _____, you would shrug off the claim that yawning is contagious. This is essentially the idea behind the null hypothesis – the is ____ treatment effect, ____ difference between groups, ____ change. In words:

The mean number of people who yawn when exposed to someone who yawns (μ_Y) is _____ the mean number of people you yawn spontaneously without seeing someone yawn (μ_N).

In symbols: H_0: _____ or H_0: _____

Alternative Hypothesis

If you were _____, you would try to justify the claim that yawning ___ contagious using evidence from the data collected. This is the idea behind the alternative hypothesis – there _____ is a treatment effect of there _____ a difference between the groups. In words:

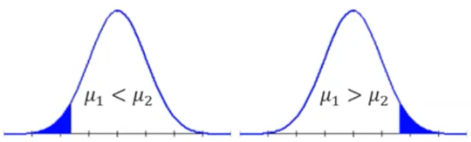
The mean number of people who yawn when exposed to someone who yawns (μ_Y) is _____ t_____ the mean number of people who yawn spontaneously without seeing someone yawn (μ_N).

In symbols: H_a: _____ or H_a: _____

Alternative Hypothesis

When writing an alternative hypothesis, there are two types:

_____ _____



A one-sided H_a is used when we would like to show _____ that one group has a _____ (_____) mean than the other. The direction is based on _____.

A two-sided H_a is used when we would like to show _____ that the groups have _____ means without regard for which is _____ or _____.

Alternative Hypothesis

In the yawning examples, we are only interested in determining if the treatment of seeing someone yawn _____ the average number of yawns.

Name_____

Alternative Hypotheses

Just as with confidence intervals, you must be aware of the _____ in which the _____ are subtracted. For the yawning example, _____ the order that the means are subtracted will necessarily _____ the inequality for a _____ H_a.

H_a: _____ or H_a: _____

Summing Up Hypotheses

For hypotheses about a _____ in means:

- The _____ is a statement of _____, typically _____
- The alternative always contains a strict _____.

  H_a: _____
  H_a: _____ } _____
  H_a: _____ – _____
- Never refer to _____ (such as $\bar{x}_1$ or $\bar{x}_1$) in the _____!
- Remember to _____ any _____ or _____ you use.

Multiple Choice Example

Three college roommates – one majoring in chemistry, one majoring in physics, and one majoring in statistics – were arguing over dinner about who had the more difficult course load that semester. The chemistry major and the physics major were quite sure they were taking the most difficult class of all, but the statistics major was doubtful (and indifferent). They decided to investigate by taking a separate random sample of words from the chemistry textbook and the physics textbook used in their most difficult courses to find the mean word length. The statistics major believed this would determine if there was a difference in the difficulty of the courses. The statistics major painstakingly obtained a random sample of 200 words from the chemistry textbook and found a mean word length of 5.71 letters with a standard deviation of 3.02 letters. A separate random sample of 200 words from the physics textbook produced a mean word length of 6.03 letters with a standard deviation of 3.58 letters. Which pair of hypotheses is appropriate for the statistics major's proposed test?

Multiple-Choice Example (Cross out incorrect answer as you watch the video!)

(A) $H_0: \mu_C = 5.71$
$H_a: \mu_p > 5.71$

(B) $H_0: \mu_P - \mu_C < 0$
$H_a: \mu_P - \mu_C > 0$

(C) $H_0: \mu_P - \mu_C = 0$
$H_a: \mu_P - \mu_C < 0$

(D) $H_0: \bar{x}_P - \bar{x}_C = 0$
$H_a: \bar{x}_P - \bar{x}_C \neq 0$

(E) $H_0: \mu_P - \mu_C = 0$
$H_a: \mu_P - \mu_C \neq 0$

What Should We Take Away?

How do we state a null hypothesis in a test for a difference in means?

H_0: _____ = 0 or H_0: _____

How do we state an alternative hypothesis in a test for a difference means?

H_a = _____ or H_a= _____
H_a = _____ or H_a= _____
H_a = _____ or H_a= _____

Name_____

AP Statistics CED 7.5 Daily Video 2 (Skill 4.C)

Setting up a Test for the Difference of Two Population Means

What Will We Learn?

How do we identify an appropriate significance test procedure for a difference in means?
How do we verify the conditions for performing a significance test for a difference in means?

Contagious Yawning

Three students wanted to test out the urban legend that yawning is contagious. From a group of 27 volunteers, they randomly assigned 14 people to have a children's nighttime store read to them while the storyteller yawned occasionally. The remaining 13 volunteers were read the same children's nighttime story, but without any yawning by the storyteller. The subjects in both groups were observed and the number of times each person yawned was recorded, as shown in the table below.

Is there convincing statistical evidence that people yawn more, on average, when watching someone yawn? Use an $\alpha = 0.05$ level of significance.

Yawn	7	3	7	6	8	6	5	6	3	5	6	7	4	5
No Yawn	3	2	2	2	0	1	2	1	1	4	3	0	5	

Identifying the Procedures

How many groups? We are comparing _____ treatment groups (_____ and _____)

What type of data were collected? Measurements are number of yawns per person (_____)

What are we asked to do? Determine if there is _____ for the claim that people yawn more, _____, when watching someone yawn than not watching someone yawn.

The procedure: _____ for a difference in population means

Check the Conditions

To check for independence:

1. The data should come from _____ samples OR a _____ experiment.
2. When sampling _____, the samples should be less than or equal to _____ of their respective populations.

To check that the shape of the _____ distribution of ($\bar{x}_1$ - $\bar{x}_2$) is approximately normal:

3. Both _____ and _____ OR both samples should come from approximately _____ distributed populations.
   - If samples are _____, the shapes of the _____ distribution should be free from _____ or _____.

Checking the Conditions (Don't forget to ✓ your conditions!)

From a group of 27 volunteers, they randomly assigned 14 people to have a children's nighttime store read to them while the storyteller yawned occasionally. The remaining 13 volunteers were read the same children's nighttime story, but without any yawning by the storyteller.

Yawn	7	3	7	6	8	6	5	6	3	5	6	7	4	5
No Yawn	3	2	2	2	0	1	2	1	1	4	3	0	5	

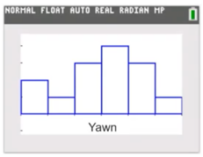
1. Both treatment groups were _____.
2. Are we sampling _____ replacement? NO! This is a _____ experiment with volunteers. This condition is _____.
3. Both samples have less than _____ so we must verify there is no extreme _____ or _____.

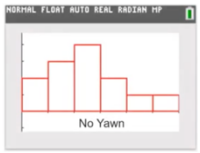
Name_____

Checking the Conditions

We will turn to the calculator to check the shape of the sample distributions:

Yawn	7	3	7	6	8	6	5	6	3	5	6	7	4	5
No Yawn	3	2	2	2	0	1	2	1	1	4	3	0	5	





Both samples are _____ and neither sample shows extreme _____ or obvious _____. It is _____ to assume they come from _____ normal populations. All conditions are _____.

How Skewed is Too Skewed?

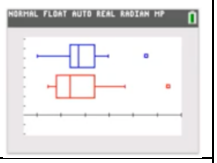
When dealing with small samples sizes, it is possible to get some unusual looking histograms, the key is to look for _____ skewness and outliers.

Bigfoot Sighting

A podiatrist noticed that female patients who wore high-heeled shoes appeared to have a higher incidence of large toe joint inflammation, or bunions, than those who did not. The podiatrist wondered if all women who regularly wear high-heeled shoes have wider feet, on average, than those who do not, indicating that high-heeled shoes may promote bunion formation. The podiatrist contacts all current female patients and asks if they were high-heeled shoes regularly (more than twice a week) or not. A random sample of 12 patients is selected from those who responded they wear high-heeled shoes regularly and a separate random sample of 12 patients is selected from those who said they do not wear high-heeled shoes regularly. The width of each patient's left foot is measured. The podiatrist wants to know if the data provide convincing statistical evidence that the mean foot width for woman who regularly wear high-heeled shoes is greater than for those who do not. Have the conditions for inference been met? (Highlight important information as your watch the video.) Remember to make a ✓!

Wears High-Heeled Shoes (cm)	9.5	10.1	9.8	9.6	9.9	9.7	9.4	10.8	9.4	9.5	9.1	8.8
Does Not Wear High Heeled Shoes (cm)	9.8	9.7	9.1	9.9	9.1	10.4	9.4	9.2	9.2	11.2	9.4	9.0

1. Both samples of female patients were _____.
2. Check 10% condition when sampling without replacement. We must assume that the podiatrist has at least _____ female patients who _____ high-heeled shoes regularly and at least _____ female patients who _____ high-heeled shoes regularly.
3. Simple samples are < 30, we must verify the sample distributions have no extreme skewness



What Should We Take Away?

How do we identify an appropriate significance test procedure for a difference in means?

Use a _____

How do we verify the conditions for performing a significance test for a difference in means?

1. The data come from _____ samples OR a _____+_____ experiment.

2. When sampling _____ replacement, the sample should be ≤ _____ of their respective populations.

3. both n_1 _____ and n_2 _____ OR they should come from _____ normal populations.

Name_____

AP Statistics CED 7.9 Daily Video 1 (Skill 3.E)

Carrying Out a Test for the Difference of Two Population Means

What Will We Learn?

How do we calculate an appropriate test statistic in a test for the difference of two population means?
How do we calculate a *p*-value in a test for the difference of two population means?

Contagious Yawning

Three students wanted to test out the urban legend that yawning is contagious. From a group of 27 volunteers, they randomly assigned 14 people to have a children's nighttime store read to them while the storyteller yawned occasionally. The remaining 13 volunteers were read the same children's nighttime story, but without any yawning by the storyteller. The subjects in both groups were observed and the number of times each person yawned was recorded, as shown in the table below. Is there convincing statistical evidence that people yawn more, on average, when watching someone yawn? Use an $\alpha = 0.05$ level of significance.

Yawn	7	3	7	6	8	6	5	6	3	5	6	7	4	5
No Yawn	3	2	2	2	0	1	2	1	1	4	3	0	5	

Recall from the previous video that the hypotheses for this significance test of a difference of means are:

H_0: _____ and H_A: _____

Where μ_Y is the _____ number of yawns per person when _____ someone yawn and μ_Y is the _____ number of yawns per person when _____ someone yawn. Additionally, the conditions for inference have _____. So, we will calculate the _____ and the _____.

Formula for Success!

All information for calculating the test statistic can be found on the AP Statistics Formula Sheet. Find each section as your watch the video. Remember:

Standardized test statistic: $\frac{statistic - parameter}{standard\ error\ of\ the\ statistic}$; and $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$ and $s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$

Calculating the Test Statistic

Let's find the summary statistics for each treatment group (use calculator):

$\bar{x}_Y$ = _____ $\bar{x}_N$ = _____ $\bar{x}_Y - \bar{x}_N$ = _____ = _____

s_Y = _____ s_N = _____

n_Y = _____ n_N = _____

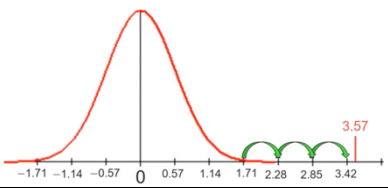
At first glance, this looks like striking difference in the _____ number of yawns per person between the two groups. This is not enough to _____ that the yawning is contagious. We need evidence in the form of a _____. How _____ is it that this experiment would produce a difference in means as _____ as _____ yawns or even _____ if yawning really is _____?

Calculating the Test Statistic approximately normal samples distribution of $(\bar{x}_Y - \bar{x}_N)$

We assume the H_0 is true: H_0: _____ = 0

From the summary statistics we can calculate and sketch the standard error:

$s_{\bar{x}_1 - \bar{x}_2}$ = _____ = _____



STATS MEDIC

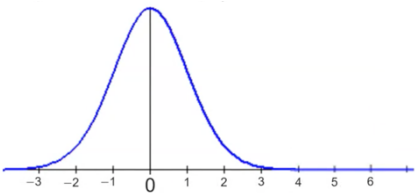
Name_____

Calculating the Test Statistic

What percent of the time would we _____ to see a difference in means of _____ yawns or something _____ extreme, assuming H_0 is true?

We will use the sampling distribution of the test statistic, which is a *t*-distribution (with df = _____ found using technology)

Using the standardized test statistic given by:



$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ = _____ = _____ (Plot this on the graph above!)

Calculating the *p*-value

What _____ of the time would we expect to see a standardized test statistic of t = _____ or something _____, assuming H_0 is _____?

Hypothesis: H_a: $\mu_Y - \mu_N > 0$ and a $P(t \geq$ _____$)$ = *p*-value = _____

Technology to the Rescue! (Follow along and perform the 2-Sample *t*-test on the calculator!)

t = _____; *p*-value = _____; df = _____

Make sure to record _____ of these values on the AP Exam!!

Multiple-Choice Practice

The Excesspresso Coffee Company wanted to appeal to customers who may want to limit their caffeine intake by advertising that their light roast coffee beans have less caffeine preserving, on average, than dark roast coffee beans. The company's head roaster used a random sample of 1`0 servings from batches of light roast coffee beans and a random sample of 10 servings from batches of dark roast coffee beans to brew a shot of expresso from each serving. Ultraviolet spectroscopy was used to determine the caffeine content inn each shot of expresso (in milligrams). Based on the summary statistics in the table below, which of the following gives the correct test statistic and *p*-value for an appropriate test that would allow the coffee company to make the advertisement described? Assume the conditions for inference have been met.

	n	Mean Caffeine Content (mg)	Standard Deviation of Caffeine Content (mg)
Light Roast Coffee	10	50.49	0.25
Dark Roast Coffee	10	50.74	0.88

Multiple-Choice Practice

State Hypotheses:

H_0: _____

H_a: _____

(A) $z = -0.25$, *p*-value = 0.63

(B) $z = -0.864$, *p*-value = 0.194

(C) $t = -0.864$, *p*-value = 0.407

(D) $t = -0.864$, *p*-value = 0.203

(E) $t = -0.25$, *p*-value = 0.203

Procedure: _____ (This is a _____ test!)

Using technology, calculate the test-statistic and the *p*-value. (Eliminate answers as you watch!)

What Should We Take Away?

How do we calculate an appropriate test statistic in a test for the difference of two population means?

We use:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where the *t*-statistic has an approximate *t* distribution with degrees of freedom found using technology.

How do we calculate a *p*-value in a test for the difference of two population means?

If H_a: $\mu_1 - \mu_2 > 0$ or $\mu_1 > \mu_2$, *p*-value = $P(t \geq$ observed test statistic)

If H_a: $\mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$, *p*-value = $P(t \leq$ observed test statistic)

If H_a: $\mu_1 - \mu_2 > 0$ or $\mu_1 > \mu_2$, *p*-value = 2 ✕ $P(t \geq |$observed test statistic)

Name_____

AP Statistics CED 7.9 Daily Video 2 (Skill 4.B)

Carrying Out a Test for the Difference of Two Population Means

What Will We Learn?

How do we interpret the *p*-value in a significance test for the difference of two population means?
How do we state a conclusion in a significance test for the difference of two population means?

Contagious Yawning

Three students wanted to test out the urban legend that yawning is contagious. From a group of 27 volunteers, they randomly assigned 14 people to have a children's nighttime store read to them while the storyteller yawned occasionally. The remaining 13 volunteers were read the same children's nighttime story, but without any yawning by the storyteller. The subjects in both groups were observed and the number of times each person yawned was recorded, as shown in the table below.

Yawn	7	3	7	6	8	6	5	6	3	5	6	7	4	5
No Yawn	3	2	2	2	0	1	2	1	1	4	3	0	5	

Is there convincing statistical evidence that yawning is contagious? That is do the data provide convincing statistical evidence that people yawn more, on average, when watching someone yawn? Use an $\alpha = 0.05$ level of significance.

Parameters & Hypotheses:

Let μ_y = _____ and

Let μ_N = _____

H_0: _____ OR H_0: _____

H_A: _____ OF H_A: _____

Procedures and Conditions:

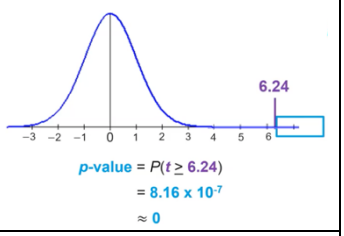
The conditions for a _____ have been verified.

Mechanics:

The test statistics is t = _____ with df = _____ and p-value ≈ _____.

Interpreting a *p*-value

Recall from the previous video that the *p*-value is calculated as the _____ of obtaining a result as _____ as the one in the study, or _____ extreme by _____ alone, assuming the _____ hypothesis is _____.



In context

Assuming _____ is true, there is nearly a _____ of getting a difference (*Yawn – No Yawn*) in mean yawns per person of _____ or _____ just by the chance involved in the _____ assignment. Getting this large for difference is _____ likely to happed just due to random assignment is there is no _____ effect, but it _____ happened!!

Stating a Conclusion: A General Guide

When we state our conclusion, it has two main parts:

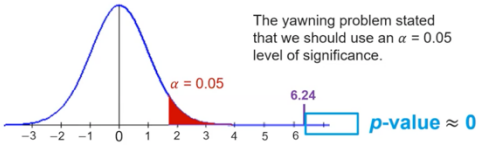
1. How does the *p*-value _____ to our level of significance, α, and what _____ must be made about H_0?
2. What does this mean about H_A in _____?

Name_____

For small *p*-values ⟹ test statistics is _____ to occur by random chance alone
Since the *p*-value _____ ≤ α = _____, we _____ H_0.
There _____ convincing _____ evidence that [state H_A in context].
For large *p*-values ⟹ test statistics is _____ to occur by random chance alone
Since the *p*-value _____ > α = _____, we _____ H_0.
There _____ convincing _____ evidence that [state H_A in context].

Stating a Conclusion

Since the _____ of close to ____ is _____
the specified _____ we reject the _____.
There _____ convincing statistical evidence to suggest _____,
that people yawn more, _____, when watching someone yawn.



Overcaffeinated?

The Excesspresso Coffee Company wanted to appeal to customers who may want to limit their caffeine intake by advertising that their light roast coffee beans have less caffeine per serving, on average, than dark roast coffee beans. The company's head roaster used a random sample of 10 servings from batches of light roast coffee beans and a random sample of 10 servings from batches of dark roast coffee beans to brew a shot of expresso from each serving. Ultraviolet spectroscopy was used to determine the caffeine content inn each shot of expresso (in milligrams). The head roaster's test for the difference in mean caffeine content of light roast and dark roast coffee beans produced a p-value of 0.203. What conclusion can the company make, at the α = 0.05 significance level, regarding their proposed advertisement?

Pause the video and write your answer here: (You might want to highlight important information?).

Let μ_L = _____
Let μ_D = _____

H_0: _____ and H_A: _____

Since the *p*-value of 0.203 _____, we _____.
There _____ convincing statistical evidence to suggest that _____
coffee beans have _____ caffeine, on average, than _____. The company should _____.

Common Mistakes

Make sure to give and _____ of the *p*-value and the significance level.
Never say that you _____, only say that you _____.
Never say that H_A is definitively _____ or _____ in a conclusion because a Type I or Type II error could have been made. We are commenting on the statistical evidence presented regarding H_A

What Should We Take Away?

How do we interpret the *p*-value in a significance test for the difference of two population means?

Assuming H_0 is true, there is a <*p-value*> probability of getting a difference in sample means of <*observed difference*> or <*greater/less/more different*>, by chance alone (or random assignment).

How do we state a conclusion in a significance test for the difference of two population means?

Since the *p*-value of ___ ≤ α = ___, we reject H_0.
There is convincing statistical evidence that [state H_a in context].

Since the *p*-value of ___ > α = ___, we fail to reject H_0.
There is not convincing statistical evidence that [state H_a in context].

STATS MEDIC

Name_____

Name_____

AP Statistics CED 7.9 Daily Video 3 (Skill 4.E)

Carrying Out a Test for the Difference of Two Population Means

What Will We Learn?

How do we perform a complete significance test for the difference of two populations means?

2007 Form B, Question 5

A serum cholesterol level above 250 milligrams per deciliter (mg/dl) of blood is a risk factor for cardiovascular disease in humans. At a medical center in St. Louis, a study to test the effectiveness of a new cholesterol-lowering drug was conducted. One hundred people with cholesterol levels between 250 mg/dl and 300 mg/dl were available for this study. Fifty people were assigned at random to each of the two treatment groups. One group received the standard cholesterol-lowering medication and the other group received the new drug. After taking the drug for three weeks, the 50 subjects who received the standard treatment had a mean decrease in cholesterol level of 10 mg/dl with a standard deviation of 8 mg/dl, and the 50 subject who received the new drug had a mean decrease of 18 mg/dl with a standard deviation of 12 mg/dl.

Does the new drug appear to be more effective than the standard treatment in lowering mean cholesterol level? Give appropriate statistical evidence to support your conclusion. (Highlight the important information in the question.)

Parameters and Hypotheses

Let μ_S = _____

and

μ_N =

_____

Hypotheses:

H_0: _____ OR H_0: _____

H_a: _____ OR H_a: _____

Procedure:

This is a

_____

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Conditions: (Be sure to ✓conditions!)

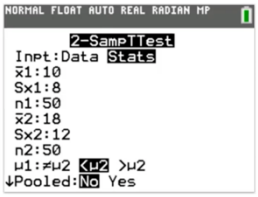
1. The subjects were _____ to either the new drug or the standard drug.
2. This is an _____, so the _____ does not apply.
3. Since _____ and _____, the sampling distribution of _____ is approximately normal (by the CLT).

The conditions are _____.

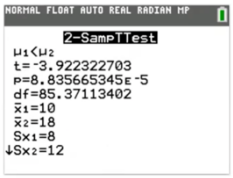
Name_____

Mechanics: (Use the calculator to calculate the *t*-statistic and the *p*-value as you watch the video.)

Input Stats



Gather Results



t = _____

p–value = _____

df = _____

Conclusion:

Since the *p*-value = _____ < _____, we _____.

There is sufficient _____ that the mean cholesterol reduction is _____ for the _____.

Scoring

Section 1: State a correct pair of

_____

Section 2: Identify the _____ or formula and _____.

Section 3: Show the correct mechanics, including the _____, _____, and _____.

Section 4: State a correct _____ using the results of the test.

What Should We Take Away?

How do we perform a complete significance test for the difference of two populations means?

Be sure to:

- ☐ **Define the _____**
- ☐ **State the _____**
  - *** Always use equality in the _____.**
  - *** Indicate the correct _____ in the _____. (<, >, ≠)**
- ☐ **Identify _____ you are using.**
- ☐ **Verify that the _____ for the procedure are _____.**
- ☐ **Calculate the _____, _____, and _____.**
- ☐ **_____ the results _____.**

Name_____

AP Statistics CED 7.10 Daily Video 1

Skills Focus – Selecting Inference Procedures

What Will We Learn?

How do we identify data that are paired?

How do we identify data that come from two samples?

What a Handful!

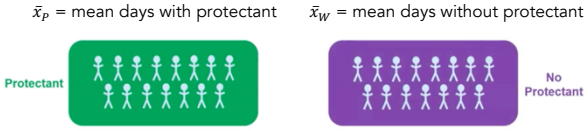
Imagine I own a leather company and I have formulated a protectant that can be applied to leather work gloves that helps them last longer by reducing the damage caused by moisture and dirt. I plan to measure the number of days the gloves endure until they are compromised by a tear or a hole. Suppose I have 30 volunteers available who use leather work gloves 5 days a week for their jobs.

What a Handful!

There are two methods I am considering to determine whether the protectant works.

1. Randomly assign 15 of the volunteers to use gloves treated with the protectant while the remaining 15 use gloves not treated with the protectant.

Notice method 1 _____ assigns the volunteers to _____.



When you proceed with an inference procedure, you will need to calculate the _____ days the gloves endured _____, then find the _____ in those means. This implies we will need to use a _____ or a _____, where _____ = true mean number of days gloves endure with protectant and _____ = true mean number of days gloves endure without the protectant.

2. For each of the 30 volunteers, randomly assign one glove (left or right) to be treated with the protectant while the opposite glove is left untreated.

Notice method 2 does _____ assign each volunteer to a distinct treatment group, but rather each volunteer undergoes _____ treatments on the gloves.

This is a matched pairs experiment!

When we proceed with an inference procedure, we will need to measure the number of days _____ gloves endured _____ and find the difference. Then we will take the _____ of those 30 differences. Each volunteer's right hand is _____ of their left hand! These are _____! This implies we will use a _____ or a _____ where _____ = _____ (*protectant – no protectant)* in the number of days gloves endure.

Helpful Hints

When _____ are provided, the observations will be given in _____. Look for a single experimental unit or a single pair to have _____ measurements given. Sometimes the _____ will also be provided.

Subject	1	2	3	4	5	6	7	8
Treatment	39.8	39.9	41.8	41.9	40.2	42	40.2	40.9
Control	40.3	40.5	40.8	40.6	40.2	40.9	40.3	40.7
Difference (T–C)	–0.5	–0.6	1.0	1.3	0	1.1	–0.1	0.2

Use a _____ or a _____ for inference about μ_d.

Helpful Hints

When data come from _____ (or two randomly assigned treatment groups), the data will be provided as _____ groups of measurements with no _____.

Group 1	15.2	11.8	11.6	5.0	8.4	18.6	15.7	22.0
Group 2	9.6	22.4	19.2	25.6	16.0	18.7	17.1	12.8

Use a _____ or _____ for inference about $\mu_1 - \mu_2$.

Name_____

2004 Form B, Question 4

The principal at Crest Middle School, which enrolls only sixth-grade students and seventh-grade students, is interested in determining how. Much time students at that school spend on homework each night. The table below shows the mean and standard deviation of the amount of time spent on homework each night (in minutes) for a random sample of 20 sixth-grade students and separate random sample of 20 seventh-grade student at this school. Based on the dotplots of these data, it is not unreasonable to assume that the distribution of times for each grade is approximately normally distributed. (Highlight key information as you watch the video!)

	Mean	Standard Deviation
Sixth-grade students	27.3	10.8
Seventh-grade students	47.0	12.4

(a) Estimate the difference in mean times spent on homework for all sixth- and seventh-grade students in this school using an interval. Be sure to interpret your interval.

Use a _____, the difference in mean times spent on homework for all sixth-grade and seventh-grade students at _____+.

2004 Form B, Question 4

(b) An assistant principal reasoned that a much narrower confidence interval could be obtained if the students were paired based on their responses; for example, pairing the sixth-grade student and the seventh-grade student with the highest number of minutes spent on homework, the sixth-grade student and seventh-grade student with the next highest number of minutes spent on homework, and so on. Is the assistant principal correct in thinking that matching the students in this way and then computing a matched-pairs confidence interval for the mean difference in time spent on homework is a better procedure that the one used in part (a)? Explain why or why not.

Pairing should occur _____ the data are collected to create pairs that are as _____ to each other as possible based on criteria that might be related to the response.

What the assistant principal is proposing is _____ because it creates an _____ _____ between responses from _____ samples.

2014, Question 5

A researcher conducted a study to investigate whether local car dealers tend to charge women more than men for same car model. Using information from the county tax collector's records, the researcher randomly selected one man and one woman from among everyone who had purchased the same model of an identically equipped care from the same dealer. The process was repeated for a total of 8 randomly selected car models. The purchase prices and the differences (*woman – man*) are shown in the table below. Summary statistics are also shown.

Car model	1	2	3	4	5	6	7	8
Women	$20,100	$17,400	$22,300	$32,500	$17,710	$21,500	$29,600	$46,300
Men	$19,580	$17,500	$21,400	$32,300	$17,720	$20,300	$28,300	$45,630
Difference	$520	-$100	$900	$200	-$10	$1,200	$1,300	$670

	Mean	Standard Deviation
Women	$25,926.25	$9,846.61
Men	$25,341.25	$9,728.60
Difference	$585.00	$530.71

Do the data provide convincing evidence that, on average, women pay more than men in the county for same car model?

Use a _____, the mean difference (*women – men*) in purchase price.

What Should We Take Away?

How do we identify data that are paired? **Look for:**

- **Data that come from a _____.**
- **Two measurements taken from a _____ individual or pairing.**
- **Inference about the _____.**

How do we identify data that come from two samples? **Look for:**

- **Data that were obtained from _____ random samples.**
- **Data that were obtained from _____ with _____ assigned treatment groups.**
- **Inference about the _____.**

Name_____

AP Statistics CED 7.10 Daily Video 2

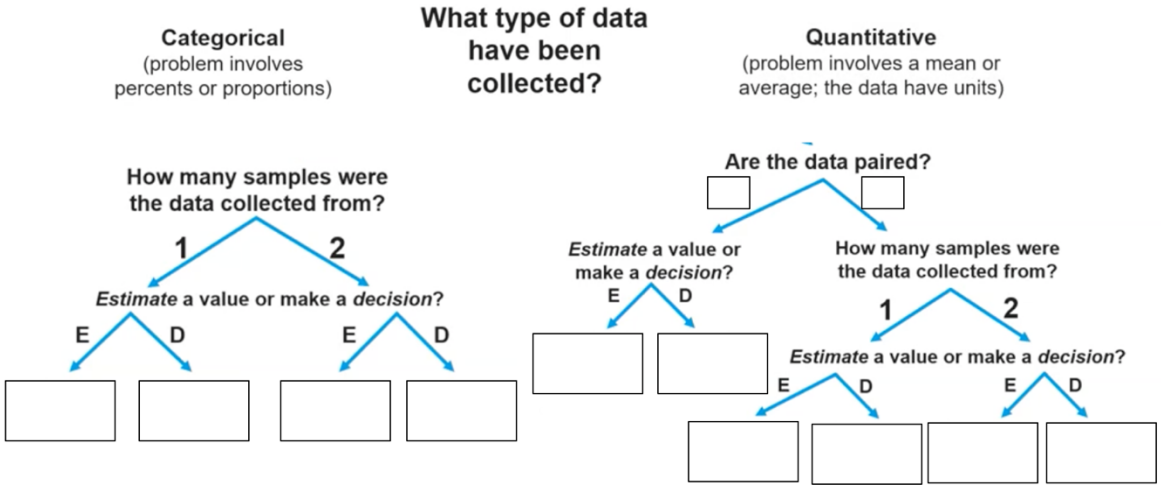
Skills Focus – Selecting Inference Procedures

What Will We Learn?

How do we identify an appropriate inference procedure involving proportions or means?

Go With the Flow (Fill in the boxes as you watch the video!)

- 1-sample *t*-interval
- 1-prop *z*-interval
- 2-sample *t*-interval
- Paired *t*-interval
- 2-prop *z*-interval
- 1-sample *t*-test
- 1-prop *z*-test
- 2-sample *t*-test
- 2-prop *z*-interval
- Paired *t*-test



Classifying Inference Problems

We will use an applet by Larry Green that allows you to practice identifying inference procedures from descriptions of studies.

http://www.ltcconline.net/greenl/java/Statistics/catStatProb/categorizingStatProblemsJavaScript.html

Name That Inference Procedure

Do people spend more time watching television than they read? 75 people were asked how many minutes they spent watching television and how many minutes they read.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name That Inference Procedure

Are sliders at Sierra-at-Tahoe more likely to spend the night in the basin than sliders at Kirkwood? 50 sierra-at-Tahoe sliders and 50 Kirkwood sliders were surveyed.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name That Inference Procedure

The average SAT score for UCLA students is 1264. You are interested to see if the average SAT score is different for UCLA students who come from rural area. 65 such students were asked what their SAT scores were.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name_____

Name That Inference Procedure

How much more productive are employees when there is music playing? A manufacturer measured the number of goods produced in a day by each of its 120 employees with and without music playing in the factory.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name That Inference Procedure

What percent of new businesses in this region are able to keep their door open a year after starting? 50 new businesses were tracked.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name That Inference Procedure

You are interested in estimating the average number of people individuals interact with each day. You survey 300 Americans.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name That Inference Procedure

Can chimps learn more words than apes? Biologists spent a year with eighty chimps and 80 apes teaching them vocabulary.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name That Inference Procedure

How effective are instant store coupons in getting customers to purchase an item? 500 customers were observed when the instant store coupon was available and 400 were observed when it was not available.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name That Inference Procedure

Does the home team win the majority of the time? 100 games were observed.

- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name That Inference Procedure

How much more effective against insects are plants with an insect resistant gene then plants without this gene? 500 plants with the gene and 500 plants without the gene were measured for the number of insects per plant

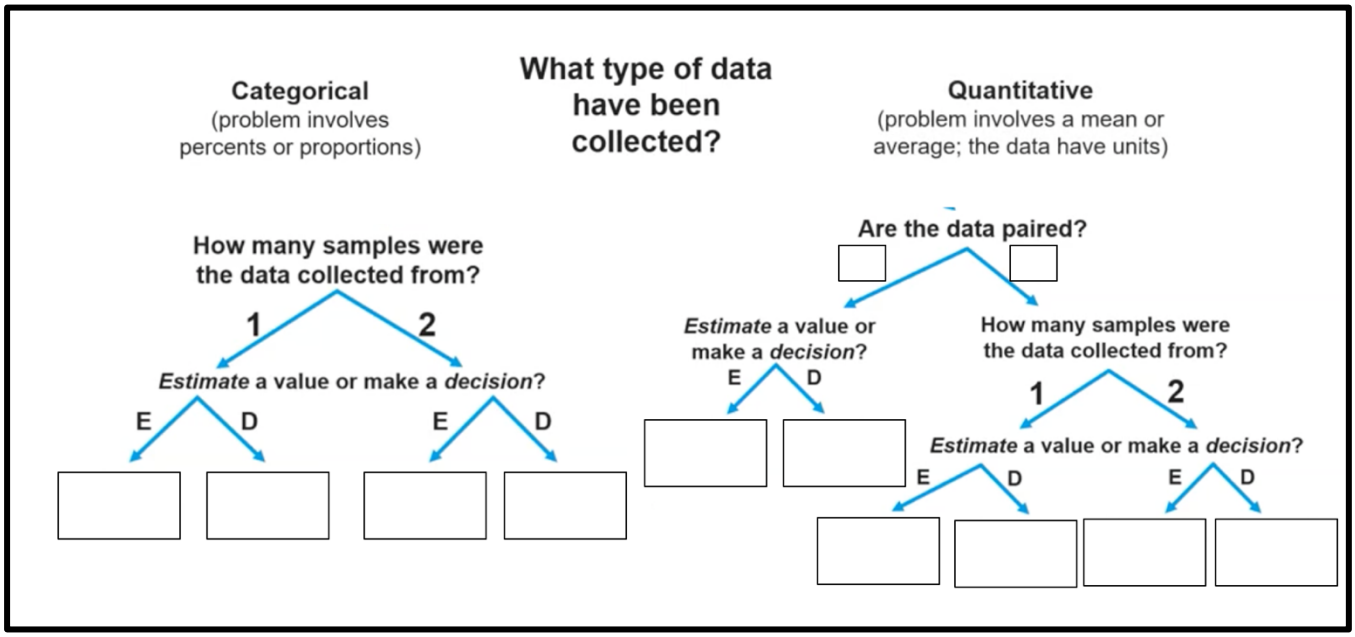
- What type of data were collected? _____
- Are the data paired? _____
- Estimate a value or make a decision? So, we use a _____

Name_____

What Should We Take Away?

How do we identify an appropriate inference procedure involving proportions or means?

- ☐ Determine if the data collected are _____ or _____.
  * For quantitative, determine if the data are _____.
- ☐ Identify _____ (or treatment groups) are being compared.
- ☐ Determine if the goal of the inference procedure is to
  _____ OR _____.



- ○ 1-sample *t*-interval
- ○ 1-prop *z*-interval
- ○ 2-sample *t*-interval
- ○ Paired *t*-interval
- ○ 2-prop *z*-interval
- ○ 1-sample *t*-test
- ○ 1-prop *z*-test
- ○ 2-sample *t*-test
- ○ 2-prop *z*-interval
- ○ Paired *t*-test