Name______________________________

AP Statistics CED 5.1 Daily Video (Skill 1.A)

Introducing Statistics – Why is My Sample Not Like Yours?

What Will We Learn?

Why do different samples produce different statistics when they come from the same population? How can we estimate the truth?

Dice

Let the random variable x represent the sum of the rolls of 2 fair dice. We would like to take a random sample of 5 rolls and find the sample mean to determine, on average, how far a player would get in 5 turns during a board game.

5 rolls: 7, 4, 9, 3, 12 with $\bar{x}$ = ____ 5 rolls: 4, 3, 10, 3, 6 with $\bar{x}$ = ___ 5 rolls 8, 2, 11, 9, 12, with $\bar{x}$ = ____

Sampling Variability

I think we can agree on a few things:

- Rolling 2 dice produces _____________ sums.
- We don't know what results we will get for ________________________________.
- The dice rolls are _________________________.
- We don't know how close our sample mean for 5 rolls is to the _______________________ mean of all possible rolls.

A Discrete Random Variable

Let the random variable X represent the sum of the rolls of 2 fair dice. The possible values of X are 2 -12.

Rolling two 1's is called snake eyes.

P(snake eyes) = ________ = _______

P(2 snake eyes) = ________ = _______

P($\bar{x}$ = 2 when n=5) ≈ _______

Die 1 Roll

Die 2 Roll	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Sample Means (Using the diagram above)

We can easily roll 2 dice 5 times and calculate the ____________________________.

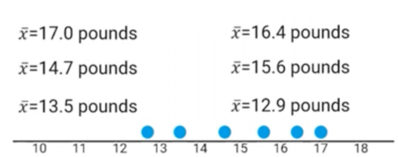
We could determine all possible _______________ for 5 rolls.

The sample means ____________________.

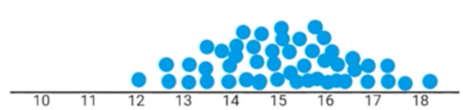
Some sample means are more ______________ than others.

Samples from Continuous RVs (random variables)

The weight of adult koalas is about 15 points with a standard deviation of 1.8 pounds. Zoo researchers tracking koala health randomly select a sample of 12 koalas each month and calculate the average weight. Each month they could get a different average weight.

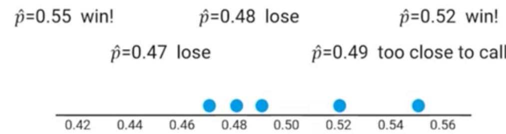


If we could continue to calculate the means of many samples of the same size from the population of all koalas, we would get a distribution of sample means that would help us draw conclusions. And notice that the sample means ___________ around the true ___________________ mean of _______.



Name________________________________

Sample Proportions



In a close election, a candidate has the support of 51% of the voters. Of course, no one knows this until after the election. Five pollsters each take a random sample of 1,000 voters to determine whether the candidate will win. They find:

The sample proportions will __________. But if we collect _____________________ sample proportions with a sample size of _______, we start to see a ______________ pattern of how the results could look.

How Do Sample Statistics Behave?

Different _____________ samples of the __________________ from the __________________________ produce statistics that ________________. A statistic from a _____________ random sample is not much different that a ________________. But, the _______________ of a statistic has a _________________ pattern of _______________ related to the __________________________________.

What Should We Take Away?

Different _____________ from the ___________________________ produce _______________ statistics. If the __________________ of a population is known, then the _______________________________ from many different ______________ samples of the ____________________ from that population is _______________________.

STATS MEDIC

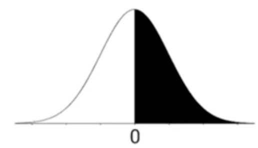
Name______________________________

AP Statistics CED 5.2 Daily Video 1 (Skill 3.A)

The Normal Distribution, Revisited

What Will We Learn?
How do we calculate the probability that a particular value lies with a given interval of a normal distribution?
How can we determine the interval associated with a given area in a normal distribution?

The Standard Normal Distribution
In Video 3 of Lesson 1.10, we learned to use z-scores to calculate the proportion of values within a given region of a normal distribution.

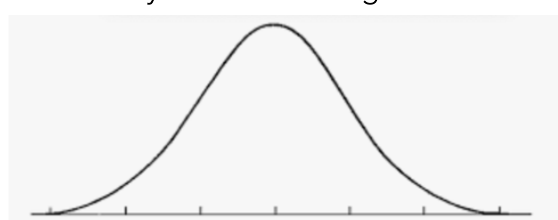


Normally Distributed Continuous RVs (random variables)
In video 1 of Lesson 4.7, we learned that continuous random variables can take on an infinite number of values in an interval on a number line. The distribution of all values may be approximately normal.

Proportion of Percent
Looking at the shaded area (above) we could say: The _____________ of values above 0 = _______.
OR The _____________ of values above 0 is 50%.

Normal Distribution Probabilities
OR, we could say that the _______________ of randomly selecting a point in the ___________ region of the given normal distribution (above) is ______________. Proportion, ______________________, ____________________ they all work and they all start with _____.

Giraffes
Giraffes are clearly the tallest living land animal partly due to their lengthy necks. The lengths of adult giraffe necks can be reasonably approximated by a normal distribution with a mean of 5.9 feet and a standard deviation of 0.3 feet. What is the probability that a randomly selected adult giraffe has a neck greater then 6 feet long?



First: define your random variable *X*.
X = ____________________________________

Second: label and shade the graph to the right with the given mean and standard deviation as described in the video.

Third: calculate the *z*-score using the formula: ____.

Finally: using technology or Table A find: $P(X > 6)$ = _____________________

What is important to include in a solution?

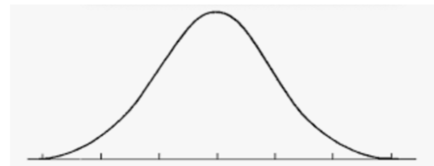
- A **defined** random variable.
- A **normal distribution** with **parameters identified.** You can show you used a **normal distribution** by: drawing a sketch as long as you have scaled ______________ standard deviations each way; using the ________________ coupled with the ___________________ statement; OR using the ____________________ statement to clearly state that you are using the normal curve. Identify parameters by simply ____________ them; including the ___________ in the z-score calculation or including ____________ in the technology statement.
- The **value** of interest. You can show this by the __________ in the sketch; by the _________ coupled with the answer or by clearly indicating the __________ and _________ bounds used in the technology statement. (Make sure you identify the values in the technology statement.)
- The **correct** probability

Name______________________________

Giraffes (Be sure to stop the video and try these problems before checking!!)

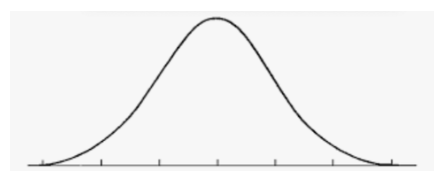
Giraffes are clearly the tallest living land animal partly due to their lengthy necks. The lengths of adult giraffe necks can be reasonably approximated by a normal distribution with a mean of 5.9 feet and a standard deviation of 0.3 feet. What is the probability that a randomly selected adult giraffe has a neck greater then 6 feet long?

(a) What is the probability that a randomly selected adult giraffe has a neck length less than 5.25 feet?



(b) What is the probability that a randomly selected adult giraffe has a neck length between 5.25 feet and 6.25 feet?

(c) What neck length would place an adult giraffe at the 25th percentile of the distribution? (Hint this is an **Inverse Normal** problem)



What Should We Take Away?

The probability a value lies within a defined interval of a ____________________________ can be calculated by finding the ___________ of the defined interval.

The _________________ of an interval associated with a given area in a normal distribution can be determined by using __________________ or ___________________.

Name________________________________

AP Statistics CED 5.2 Daily Video 2 (Skill 3.A)

The Normal Distribution, Revisited

What Will We Learn?

When can a linear combination of two random variable be modeled by a normal distribution?
How do we calculate a probability involving such a linear combination?

The Race

The Capitol 10K is a fun run in Austin, Texas, complete with live music along the route, some runners in costumes, but also some very competitive runners. The wide variety of fitness levels creates a large standard deviation of tunning times for the close to 20,000 participants.

Run Time Parameters

Females	Males
X = completion time (min) for a randomly selected female.	Y = completion time (min) for a randomly selected male.
$\mu_X = 81 \quad \sigma_X = 15.7$	$\mu_Y = 70 \quad \sigma_Y = 13.3$

The completion times for each gender are _____________ and may be reasonably approximated by a ____________ distribution. The time it takes for each runner to complete the course is measured by a tracking device and recorded for gender.

What is the probability that a randomly selected female finishes the race faster than a randomly selected male? (Note: the mean time for males to finish the race is ___________; for females it takes them ___________ longer.)

First, we need to find the parameters of the distribution of _____________ in completion times (*X* – *Y*). Remember that is (*X* – *Y*) or (*Females* – *Males*). Picture randomly selecting a ________ then randomly selecting a male and then _____________ their race times. And repeating this process for ALL possible combinations of females and males out of the 20,000 racers.

To find the _________ of the differences, we would just take the ______________ of the ___________.

$\mu_{X-Y} = \mu_X - \mu_Y = 81 - 70 = 11$ Thus, on average the males finish the race ___________ faster than the females.

To find the standard deviation of the differences you cannot ___________ the standard deviations! You must first change them to _____________ and then we always ________.

Because *X* and *Y* are _________________ you have: $\sigma_{X-Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{15.7^2 + 13.3^2} = 20.576$

Shape of the Distribution of *X* – *Y*.

$$\mu_{X-Y} = 11 \qquad \sigma_{X-Y} = 20.576$$

Still considering the question: What is the probability that randomly selected female finishes the race faster than a randomly selected male?

We know that shape of the distributions of *X* – *Y* is _____________________________________. The distribution of the __________________ of two _______________________ random variables can be modeled with a ________________________________ if the probability distribution of each random variable is _______________________________________.

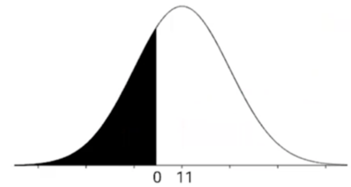
Note: In the beginning of this problem, we were **given** that the male and female race time were approximately normal. So, we can assume that the distribution of the __________________ is also ______________________________________.

Name_______________________________

No Difference

What is the probability that a randomly selected female finishes the race faster than a randomly selected male?

$\mu_{X-Y} = 11 \qquad \sigma_{X-Y} = 20.576$



Using the diagram to the right:

If the randomly selected male and female finish with the __________ time, the difference would be _______.

Understanding the Question

If the female finishes ___________, then her time would be __________, making the difference (female-male) ___________. So, what we are looking for here is the probability the difference in race time (female-male) is less than _______ or a ________________________.

Calculating the Probability

What is the probability that a randomly selected female finishes the race faster than a randomly selected male?

$\mu_{X-Y} = 11 \qquad \sigma_{X-Y} = 20.576$

First, calculate the *z*-score:

Second, using Table A or technology, find P($X - Y < 0$) =

Finally, write your answer as a complete sentence.

Shape of the Distribution of $X + Y$

$\mu_{X+Y} = \mu_X + \mu_Y \qquad \sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$

The distribution of the ______ of two independent random variables can be modeled with a ____________ distribution if the probability distribution of each random variable is _______________________________.

Another 10K

A more competitive 10K has the mean and standard deviations listed to the right for male and female finishing time in mintues. Times for males and females are independent. It is reasonable to assume that the distribution of times for each gender is approximately normal.

Females	Males
X = completion time (min) for a randomly selected female.	Y = completion time (min) for a randomly selected male.
$\mu_X = 63 \quad \sigma_X = 4$	$\mu_Y = 54 \quad \sigma_Y = 3.3$

What is the probability that a randomly selected female will finish the race at least 15 minutes slower that a randomly selectect male? (Pause the video and complete the problem, then check answer!)

What Should We Take Away?

A linear combination (including a difference or a sum) of two _______________ approximately normal random variables can be modeled by a ____.

The procedures learned to find probabilities using a normal distribution also apply when finding probabilities involving a linear combination of ________________________, approximately _______________ random variables.

Name_______________________________

AP Statistics CED 5.2 Daily Video 3 (Skill 3.C)

The Normal Distribution, Revisited

What Will We Learn?

How can we decide whether a normal distribution is a good approximation for an unknown distribution?

What characteristics of a normal distribution should we check when assessing normality?

Adulting

Although a majority of parents disagree, 14% of young adults (18-29) claim their parents are doing too little for them.

What might the results of a random sample of 10 young adults look like on this issue?

Could we use the normal distribution as an approximation for the number of young adults in a random sample of size *n* = 10 who think their parents are doing too little for them?

The Sample Count

14% of young adults claim their parents are doing too little for them. Let *X* = the number of young adults in a random sample of 10 who think their parents do too little for them.

This is a binomial situation. If it is binomial there has got to be ________ outcomes. The number of trials if _________ at ten. What one young adult says should be ______________________ of what another young adult says. And, the probability a young adult claims their parent is doing too little is 14% for _______________ trial.

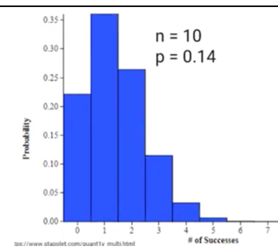
Therefore, this is a ____________ situation.

The Binomial Formula

For instance, *X* could be 3. Therefore, $P(X=3)=\binom{10}{3}0.14^3(1-0.14)^7=0.115$

We could then make a graph of ___________________ values of *X* and their probabilities to determine whether a normal distribution is an appropriate model.

Is Normal Appropriate?



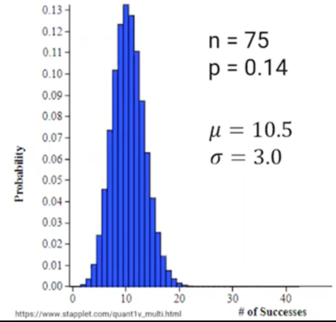
This distribution is _______________ but not ________________ or __________________________.

On the horizontal axis we have _________ possible values of ______. And on the vertical axis we have the __________________ for each one.

The distribution is clearly ____________________________, which makes normal distribution inappropriate as an approximation.

Name______________________________

Is Normal Appropriate?



Now, if we do that same thing with a samples size of _______.

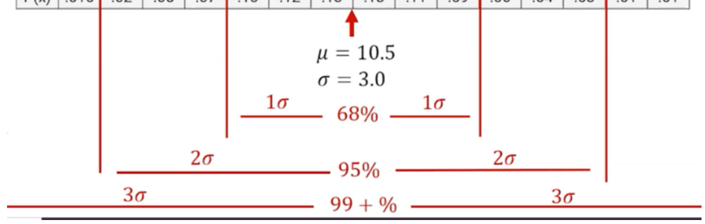
This distribution is ____________ and looks roughly _____________ and _____________________.

But is it appropriate to use normal distribution to approximate probabilities?

Partial Probability Distribution

X = the number of young adults in a random sample of 75 who think their parents do too little for them.

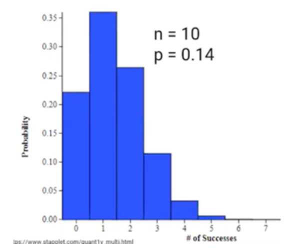
x	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
P(x)	.010	.02	.05	.07	.10	.12	.13	.13	.11	.09	.06	.04	.03	.01	.01



Note how this partial probability distribution fit the 68-95-99.7 rule for a normal distribution.

When the sample size is _______ enough, the ___________ distribution is a good approximation to a binomial distribution.

Assessing Normality



This distribution is ______________ but not _______________ or _________________________.

When the sample size is __________, the normal distribution is ______ a good approximation to a binomial distribution.

What Should We Take Away?

Characteristics to check for an approximately normal distribution:

1. Unimodal, roughly symmetric, bell-shaped
2. Empirical rule generally applies (68, 95, 99.7)

Name______________________________

AP Statistics CED 5.3 Daily Video 1 (Skill 3.C)

The Central Limit Theorem

What Will We Learn?

What is a sampling distribution?
How can we use simulation to approximate the sampling distribution of a statistic?
What does the central limit theorem say?

Sampling Distribution Definition

A sampling distribution of a statistic is the distribution of values for the statistics for ______________ __________ of the ____________________ from a given population.

You are interested in the average resting heart rate of teenagers. You use a sample size of _________. You take a ___________ sample of _____ teens. Determine the sample _________ resting heart rate. And then do it again, and again and again. Until you have ________ possible samples of 28 and their sample ________ resting heart rate. The collection of _________ those sample means is a sampling distribution. That is nearly impossible to do for most populations

Rice University Applet http://onlinestatbook.com/stat_sim/sampling_dist/

Follow along as the video demonstrates this sampling distribution applet.

If the population distribution is approximately ___________, it appears the sampling distribution is approximately ___________ even with a _____________ sample.

If the population distribution is skewed and the sample size is small, the sampling distribution is also ________________________. However, if the sample size is large enough, the sampling distribution is __________, _____________ and appears _________________________.

It appears that no matter what the population looks like, or one single sample, the ___________________________________ of sample means will be approximately normal **IF** the sample size is _________________________.

What Should We Take Away?

The sampling distribution of a statistic is the distribution of values for the statistic of _____ possible samples of the __________________ from a given population.

The sampling distribution of a statistic can be ________________ by generating repeated random samples from a population.

The central limit theorem __________ state that when the sample size is _______________________ ______________, a sampling distribution of the __________ of a random variable will be approximately ___________________ distributed.

Name______________________________

AP Statistics CED 5.3 Daily Video 2 (Skill 3.C)

The Central Limit Theorem

What Will We Learn?

What is a randomization distribution?

What can a randomization distribution tell us about the results of an experiment?

The Experiment

Newborns often sleep more during the day than the night. In a double-blind experiment to increase nighttime sleep for newborns, 16 nursing mothers were randomly assigned to take either melatonin or a placebo.

Experimental Result

During the night, newborns wake up to be fed and then, hopefully, return to sleep. The total number of hours each newborn slept for the right was recorded.

Melatonin				Placebo			
11.1	10.4	8.8	10.4	7.2	10.8	9.3	10.6
10.7	13.0	11.4	12.8	9.4	8.0	10.1	9.6

$\bar{x}_M$ = 11.07 hours $\bar{x}_P$ = 9.38 hours

$$\bar{x}_M - \bar{x}_P = 11.07 - 9.38 = 1.69 \text{ hours}$$

Randomization Distribution

To investigate whether the difference in sleep time could reasonably be due to chance, we will ________________ a randomization distribution. A randomization distribution is a collection of statistics generated by _____________ reassigning the response values to the treatment group.

Random Reallocation

If you take all of the values out of their treatment group and ______________ assign eight of the values in the melatonin group the remaining values will be in the placebo group.

11.1	10.4	8.8	10.4	7.2	10.8	9.3	10.6
10.7	13.0	11.4	12.8	9.4	8.0	10.1	9.6

Calculating Statistics

Once they are separated, with total disregard for where they really came from, calculate the __________ of each treatment group and find the _________________ in those means.

Melatonin				Placebo			
13.0	8.8	11.4	12.8	11.1	10.7	10.4	10.4
10.8	8.0	10.1	9.6	7.2	9.4	9.3	10.6

$\bar{x}_M$ = 10.56 hours $\bar{x}_P$ = 9.89 hours

$$\bar{x}_M - \bar{x}_P = 10.56 - 9.89 = 0.67 \text{hours}$$

Find the difference in mean sleep time. Here (because our value is positive) it looks like the melatonin group got _________ sleep on average then the placebo group.

Another Random Reallocation

Randomly assign another 8 values for the melatonin group. The remaining vales will be in the placebo group.

11.1	10.4	8.8	10.4	7.2	10.8	9.3	10.6
10.7	13.0	11.4	12.8	9.4	8.0	10.1	9.6

Calculate Statistics

Melatonin				Placebo			
10.4	10.7	13.0	11.4	11.1	10.4	8.8	12.8
7.2	8.0	10.6	9.6	9.4	10.8	10.1	9.3

$\bar{x}_M$ = 10.11 $\bar{x}_P$ = 10.33

$$\bar{x}_M - \bar{x}_P = 10.11 - 10.33 = -0.22$$

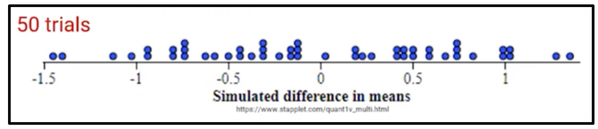
Calculate the mean of each treatment group and find the difference in means.

Here (because the value is negative) it looks like the placebo group got ________ sleep on average than the melatonin group.

Name______________________________

Random Assignment Possibilities

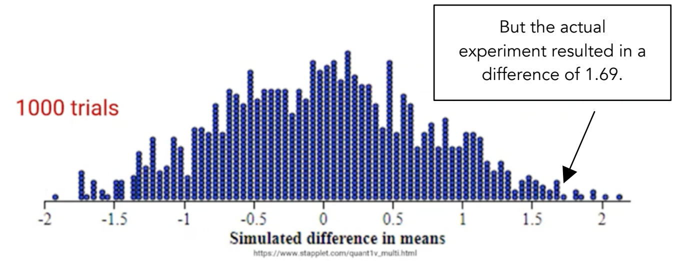
If we continue to ______________________________ the response values to the treatment groups, find the _______________ of sample means, and ____________ the differences, we will create a ____.



This dotplot represents a ________________ difference in means. (Note that each point (dot) represents the times from eight babies sleeping in the _______________ group and eight babies sleeping the ________________ group, finding the __________ of each and _______________ them.) This is what the distribution looks like for 50 trials. *Note: There are ______________ different possibilities for reallocating the treatments!!!

Random Assignment Possibilities

Using this simulated data we could find the probability that we would get ____________ 1.69 or greater hours of sleep by chance alone.



It looks like there are $\frac{12}{1000}$ = _______

Conclusion

$\bar{x}_M$ = __________________________________

$\bar{x}_P$ = __________________________________

$P(\bar{x}_M - \bar{x}_P \geq$ _______ | ____) = ________ = ________

The probability of the _______________________ for the melatonin group being ________________ _________ greater than the ______________________ for the placebo group in a ________________ __________________ of values to treatment groups is ___________. So it is ______________ that the ___________________________ happened by _____________________. This experiment provides ____________________ evidence that the melatonin given to nursing mothers ___________________ increases the _________ number of hours that newborns sleep.

What Should We Take Away?

A _________________ distribution for a randomized experiment is a collection of statistics generated by _________________ to _________________ response values to treatment groups.

A randomized distribution can be used to assess the _________________ of an observed outcome happening by _____________________.

Name________________________________

AP Statistics CED 5.4 Daily Video 1 (Skill 4.B)

Biased and Unbiased Point Estimates

What Will We Learn?

What is a point estimate?

How do we determine if an estimator is unbiased?

Weimaraner

A breeder has 5 Weimaraner dogs. A sample of 3 dogs will be randomly selected for a study. The ages of the dogs (in years) are: 0 2 5 8 10

Is the sample mean an unbiased estimator of the mean age of the population?

Unbiased Estimator. Population ages: 0 2 5 8 10 $\mu = 5$

When estimating a parameter, an estimator is ______________ if,__________________, the value of the estimator is __________ to the population parameter.

The population is small enough that we can actually investigate all possible samples of size 3.

One Sample Mean Population ages: 0 2 5 8 10 $\mu = 5$

One possible sample of three dog ages is: 0, 2, 5 $\bar{x} = 2.3$

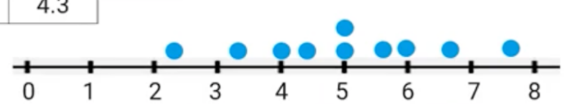
A sample statistic is a ____________________ of the corresponding population parameter

Then continue to find the sample mean for all possible samples of size 3.

All Possible Sample Means

Sample	$\bar{x}$	Sample	$\bar{x}$	Sample	$\bar{x}$
0, 2, 5	2.3	0, 5, 10	5.0	2, 5, 10	5.7
0, 2, 8	3.3	0, 8, 10	6.0	2, 8, 10	6.7
0, 2, 10	4.0	2, 5, 8	5.0	5, 8, 10	7.7
0, 5, 8	4.3				

$\mu_{\bar{x}} = 5$



When estimating a ________________ parameter, an _______________ exhibits _________________that can be modeled using probability.

A Second Look Population ages: 0 2 5 8 10 $\mu = 5$

$\mu_{\bar{x}} = 5$

Let's look at the definition of an unbiased estimator again.

Unbiased Estimator – Mean

An estimator is unbiased if ______________________, the value of the ____________________ is ______________ to the ____. Here, with the mean, it is!

One Sample Range Population ages: 0 2 5 8 10 population range = 10

One possible sample of three dog ages: 0, 2, 5 range = 5

point **estimate**

point **estimator**

Is the sample range an unbiased estimator of the population range?

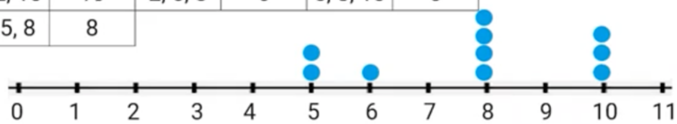
Continue to find the sample range for all possible samples of size three.

Name______________________________

All Possible Sample Ranges: Population ages: 0 2 5 8 10 population range = 10

Sample	Range	Sample	Range	Sample	Range
0, 2, 5	5	0, 5, 10	10	2, 5, 10	8
0, 2, 8	8	0, 8, 10	10	2, 8, 10	8
0, 2, 10	10	2, 5, 8	6	5, 8, 10	5
0, 5, 8	8				

mean of sample ranges = 7.8



Notice 7.8 is not equal to 10!

Notice that the different values of the range for different samples of size 3 vary, but they do NOT vary on either side of the population range. They vary only on the ___________ side!

Biased Estimator – Range

An estimator is unbiased if, ________________, the value of the _______________ is equal to the ____. Here _______________ the value of the _________________ does not equal the population parameter. So, range is a _______________ estimator of the _____________________ range.

What Should We Take Away?

A sample statistic is a _________________________of the corresponding population parameter.

When estimating a _________________________________, an estimator exhibits _________________ that can be modeled using probability.

An estimator is __________________ if, ___________________, the value of the estimator is _____________ to the population parameter.

population
range = 10

STATS MEDIC

Name________________________________

AP Statistics CED 5.5 Daily Video 1 (Skill 3.B)

Sampling Distributions for Sample Proportions

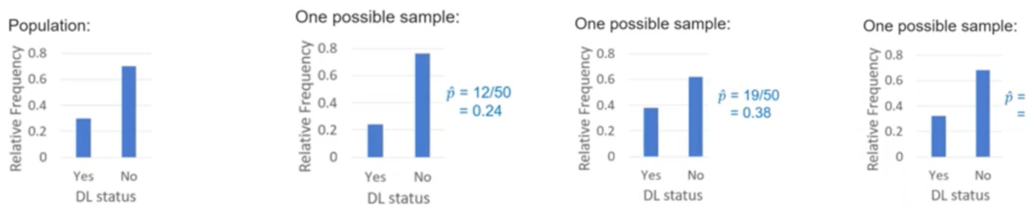
What Will We Learn?

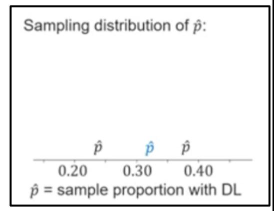
How do we determine the parameters of a sampling distribution of a sample proportion?

How do we determine if the shape of a sampling distribution of a sample proportion is approximately normal?

Sampling From a High School

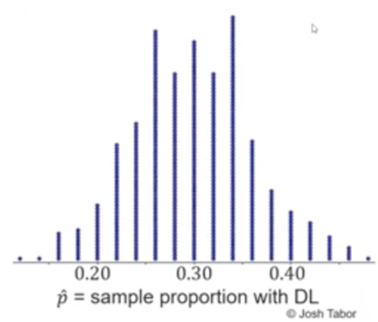
In a high school with 2000 students, 30% of the students have a driver's license. Suppose we select a random sample of *n* = 50 students from the high school, ask each student in the sample if he or she has a driver's license, and calculate the sample proportion $\hat{p}$ with a driver's license.





We won't always get the same value when we take a sample, that is called ______________________ ____________________. There are lots of possible samples of size 50 that we could take.

Here is a simulated sampling distribution of _______ = the sample proportion of students with a driver's license, based on 500 samples of size *n* = 50. When you think about describing this distribution you should always think about _____________ and ______________ and _________________.



The distribution of the sample proportion is:

- ____
- ____
- ____

Mean and Standard Deviation

Let $\hat{p}$ be the sample proportion of ________________ in a random sample of size *n* selected from a population with proportion of successes p. Then,

$$\mu_{\hat{p}} = p$$

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

*assuming the sample size is less than _________ of the population size OR if we sample with replacement from the population. This condition is met in almost all circumstances. Good news! These are on the AP Stat formula sheet!

In our example, $\hat{p}$ is the _____________ proportion of students with a driver's license in samples of size *n* = _____ from a population where ____________. So that tells us that:

the mean of the $\hat{p}$ distribution in that context is: $\mu_{\hat{p}} = p = 0.3$

and the standard deviation is:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3(1-0.3)}{50}} = 0.065^*$$

In this case we are taking samples of size _____ from a large high school of 2000 student clearly < ___________ of the population.

STATS MEDIC

Name________________________________

Shape

Let _______ be the sample proportion of success in a random sample of size *n* selected from a population with a proportion of successes ________. Then, the sample distribution of _______ will be approximately normal when: $np \geq 10 \text{ and } n(1-p) \geq 10$

For our example about students with driver's licenses, ____________ and __________ :

np = _____________________________ and

$n(1-p)$ = ___________________________________

What Should We Take Away?

How do we determine the parameters of a sampling distribution of a sample proportion?

Mean: _____________________

Standard Deviation : _________________________________,

***assuming the sample size is ___________ than ____________ of the population size.**

How do we determine if the shape of a sampling distribution of a sample proportion is approximately normal?

The sampling distribution of _______ will be approximately normal when:

_______________________ **and** _________________________________

Name________________________________

AP Statistics CED 5.5 Daily Video 2 (Skill 4.B)

Sampling Distributions for Sample Proportions

What Will We Learn?

How do we interpret the parameters of a sampling distribution of a sample proportion?

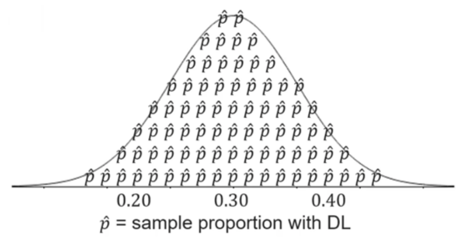
How do we calculate and interpret probabilities involving a sampling distribution of a sample proportion?

Sampling From a High School

In a high school with 2000 students, 30% of the students have a driver's license. Suppose we select a random sample of n = 50 students from the high school, ask each student in the sample if he or she has a driver's license, and calculate the sample proportion $\hat{p}$ with a driver's license.

Here is what we learned from the 5.5 video 1. The sampling distribution of $\hat{p}$ is:

- $\hat{p}$_____________________________
- $u_{\hat{p}}$ = ________ = _____________
- $\sigma_{\hat{p}}$ = ____________ = ____________ = ________



Interpreting the Mean

Interpret the value of $\mu_{\hat{p}}$ = .30

For all ____________ samples of size ______________ from this population, the sample proportions of students who have a driver's license will have a ____________ of __________.

Interpreting the Standard Deviation

Interpret the value $\sigma_{\hat{p}}$ = 0.065

For all ___________ samples of size ______________ from this population, the sample proportions of students who have a driver's license _____________________________ by about _______________ from the _____________________ proportion of _____________.

*Knowing how much a ______________ statistics ________________ varies from the ___________ is one of the most important reasons to study ____.

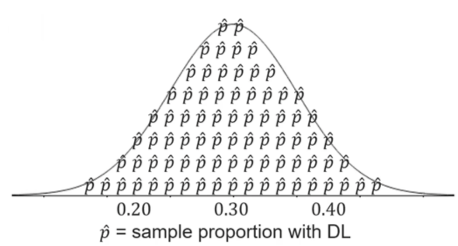
Calculating Probabilities

Would it be unusual to get a sample proportion of $\hat{p}$ = 0.12 or less in a random sample of size n = 50 from this population? (Using the normal curve below to label and shade.)

Z = ________________________ = _________________

Use table A or technology to calculate:

$P(\hat{p} < 0.12)$ = ________



STATS MEDIC

Name_______________________________

Interpreting Probabilities

Would it be unusual to get a sample proportion of $\hat{p}$ = 0.12 or less in a random sample of size n = 50 from this population?

Getting a _______________ proportion of ____________ or ____________ happens in only about ____________ of all possible samples of ________________ from _______ population. This is __________________________.

*Knowing what values of a statistic count as ______________ is one of the most important reasons to study _____________________________________.

What Should We Take Away?

How do we interpret the parameters of a sampling distribution of a sample proportion?

- In __________________
- Reference ___________ possible samples ____________________ from the population
- For standard deviation, include __________________ or _______________________.

How do we calculate and interpret probabilities involving a sampling distribution of a sample proportion?

- Use ____________________________________ procedures from Topic 5.2 with the parameters of the sampling distribution of _________ (Topic 5.5, Video 1)
- Interpretations should be ________________________ and reference ____________________ samples of ______________________ from the population.

Name________________________________

AP Statistics CED 5.6 Daily Video 1 (Skill 3.B)

Sampling Distributions for Differences in Sample Proportions

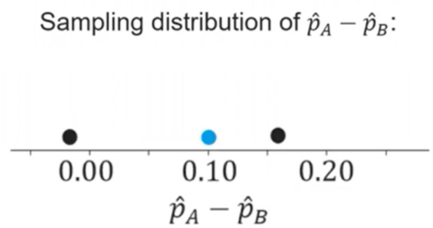
What Will We Learn?

How do we determine the parameters of a sampling distribution of a difference in sample proportions?

How do we determine if the shape of a sampling distribution of a difference in sample proportions is approximately normal?

Sampling From a High School

In high school A, 30% of the students have a driver's license. In high school B, 22% of the students have a driver's license. Suppose we select a random sample of *n* = 50 students from each high school, ask each student in each sample if he or she has a driver's license, calculate the sample proportion with a driver's license in each sample, and compute the difference in sample proportions $\hat{p}_A - \hat{p}_B$.



One possible sample from A: $\hat{p}_A$ = 18/50 = 0.36

One possible sample from B: $\hat{p}_B$ = 10/50 = 0.20

$\hat{p}_A - \hat{p}_B$ = ______________________________

One possible sample from A: $\hat{p}_A$ = 13/50 = 0.26

One possible sample from B: $\hat{p}_B$ = 14/50 = 0.28

$\hat{p}_A - \hat{p}_B$ = ______________________________

One possible sample from A: $\hat{p}_A$ = 14/50 = 0.28

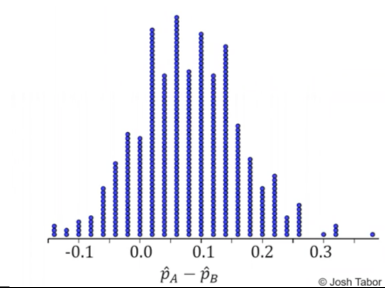
One possible sample from B: $\hat{p}_B$ = 9/50 = 0.18

$\hat{p}_A - \hat{p}_B$ = ______________________________

Here is a simulated sampling distribution of $\hat{p}_A - \hat{p}_B$ based on 500 random samples of size *n* = 50 from high school A and 500 random samples of size *n* = 50 from high school B.

The distribution of $\hat{p}_A - \hat{p}_B$ is:

- ____
- ____
- ____



Mean and Standard Deviation

Let ________ be the sample proportion of successes in a _______________ sample of size, ________ selected from population 1 with proportions of successes _______ and let ________ be the sample proportion of successes in a random sample of size ________ selected from population 2 with proportion of successes ________. Then,

$$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2 \qquad \sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Assuming the samples sizes are less than 10% of the population sizes and that the two samples are independent.

Name________________________________

Mean and Standard Deviation

In high school A, $\hat{p}_A$ = _________ and n_A = _________.
In high school B, $\hat{p}_B$ = _________ and n_B = _________.
Therefore,

$\mu_{\hat{p}_A - \hat{p}_B}$ = ________________________________

$$\sigma_{\hat{p}_A - \hat{p}_B} = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

= ___________________________________ = __________

Assuming the sample sizes are less than _____________________________________ and the two samples are _______________________.

Mean and Standard Deviation

Good news! Both of the formulas above can be found on the AP Statistics formula sheet!!! Take a moment to locate these formulas now!

Shape

Let $\hat{p}_1$ be the sample proportion of successes in a random sample of size n_1 selected from population 1 with proportion of successes p_1 and let $\hat{p}_2$ be the sample proportion of successes in a random sample of size n_2 selected from population 2 with proportion of successes p_2. Then, the sampling distribution of $\hat{p}_1 - \hat{p}_2$ will be approximately normal when:

__________ ≥ _____ ; __________ ≥ _____ ; __________ ≥ _____ ; __________ ≥ _____

For our example about students with driver's licenses from high school A and high school B,
n_A = _________; p_A = _________; n_B = _________; and p_B = _________

Check conditions: ____________________ ≥ 10; ____________________ ≥ 10
____________________ ≥ 10; ____________________ ≥ 10

Therefore the (sampling) distribution of $\hat{p}_A - \hat{p}_B$ is ____________________ normal in this context.

What Should We Take Away?

How do we determine the parameters of a sampling distribution of a difference of sample proportions? **MEAN:** $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$ **SD:** $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

Assuming the sample sizes are less than 10% of the population sizes and samples are independent.

How do we determine if the shape of a sampling distribution of a difference in sample proportions is approximately normal?

The sampling distribution of $\hat{p}_1 - \hat{p}_2$ will be approximately normal when:

$$n_1p_1 \geq 10, n_1(1-p_1) \geq 10, n_2p_2 \geq 10, \text{ and } n_2(1-p_2) \geq 10$$

Name________________________________

AP Statistics CED 5.6 Daily Video 2 (Skill 4.B)

Sampling Distributions for Differences in Sample Proportions

What Will We Learn?

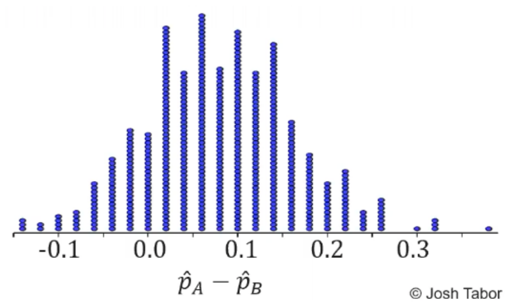
How do we interpret the parameters of a sampling distribution of a difference in sample proportions? How do we calculate and interpret probabilities involving a sampling distribution of a difference in sample proportions?

Sampling From a High School

In high school A, 30% of the students have a driver's license. In high school B, 22% of the students have a driver's license. Suppose we select a random sample of *n* = 50 students from each high school, ask each student in each sample if he or she has a driver's license, calculate the sample proportion with a driver's license in each sample, and compute the difference in sample proportions $\hat{p}_A - \hat{p}_B$.

The distribution of $\hat{p}_A - \hat{p}_B$ is

- Approximately normal
- $\mu_{\hat{p}_A - \hat{p}_B} = 0.30 - 0.22 = 0.08$
- $\sigma_{\hat{p}_A - \hat{p}_B} = \sqrt{\frac{0.30(1-0.30)}{50} + \frac{0.22(1-0.78)}{50}} = 0.087$



Interpreting the Mean: Interpret the value $\mu_{\hat{p}_A - \hat{p}_B} = 0.08$.

"For all ________________ samples of ______ students from high school A and ______ students from high school B, the ____________________ (A – B) in sample proportions of students __________________ _____________________ will have a __________ of 0.08." (See graph above)

*Note: Always be clear what order you are subtracting in.

Interpreting the Standard Deviation: Interpret the value $\sigma_{\hat{p}_A - \hat{p}_B} = 0.087$

"For all _______________ samples of size ______ students from high school A and ______ students from high school B, the _____________________ (A – B) in sample proportions of students who ____ typically vary by about ________ from the _________ difference of 0.08." (See graph above)

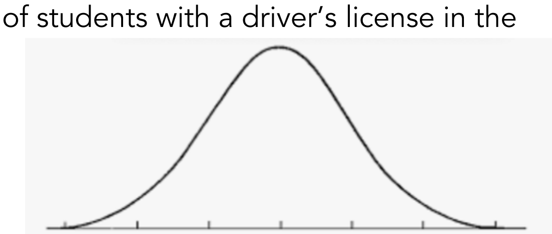
*Note: Knowing how much a sample statistic ______________ varies from the truth is one of the _______________ important reasons to study sampling distributions.

Calculating Probabilities

In high school A, 30% of the students have a driver's license. In high school B, 22% of the students have a driver's license. Suppose we select a random sample of *n* = 50 students from each high school, would it be unusual to get a greater proportion of students with a driver's license in the sample from school B?

We want to find: _______________________________

Use the blank normal curve at the right to sketch a picture then calculate the z-score:



z = ________________________. Using technology: $P(\hat{p}_A - \hat{p}_B < 0) =$ ____________________

Name______________________________

Interpreting Probabilities:
Would it be unusual to get a greater proportion of students with a driver's license in the sample from school B?
"Getting a difference (A – B) in sample proportions of ____________ happens in about ________ of all possible samples of size 50 from these populations. This is not unusual.
*Note: Knowing what values of a statistic count as unusual is one of the most important reasons to study sampling distributions.

What Should We Take Away?
How do we interpret the parameters of a sampling distribution of a difference in sample proportions?

- **In ________________**
- **Reference _____________________________ of these ___________ from the populations.**
- **For standard deviation, include "______________" or "_________________"**

How do we calculate and interpret probabilities involving a sampling distribution of a difference in sample proportions?

- **Use normal distribution procedures from Topic 5.2 with the _________________ of the sampling distributions of $\hat{p}_A$ - $\hat{p}_B$ (Video 1)**
- **Interpretations should in ____________________ and reference ______________________ samples of ___________________ from the populations.**

Name________________________________

AP Statistics CED 5.7 Daily Video 1 (Skill 3.B)

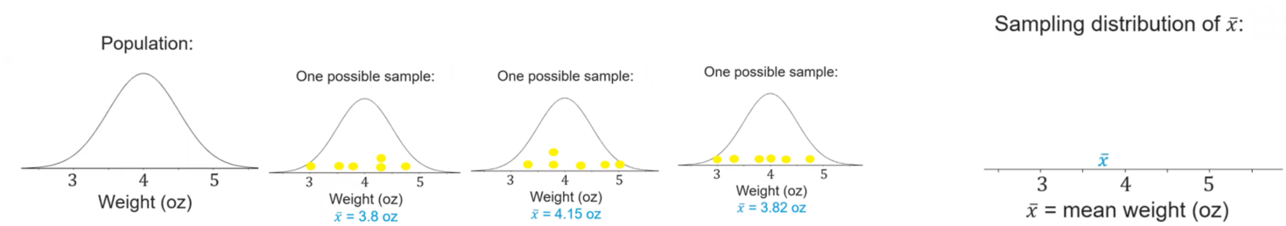
Sampling Distributions for Sample Means

What Will We Learn?

How do we determine the parameters of a sampling distribution of a sample mean?
How do we determine if the shape of a sampling distribution of a sample mean is approximately normal?

Sampling Lemons:

A large tree produces lemons with weights that are approximately normally distributed with a mean of 4 ounces and a standard deviation of 0.5 ounces. Suppose we select a random sample of *n* = 6 lemons from the tree, weigh each lemon, and calculate the sample mean weight $\bar{x}$ for each sample.

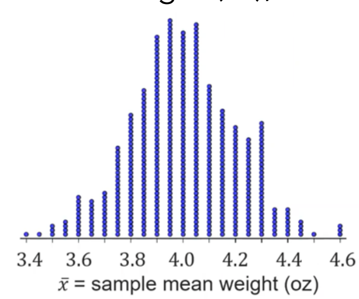


Add the $\bar{x}$ from the second and third sample to the sampling distribution to the right.

Remember, this is just three possible samples of size six from the population. There are MANY, MANY more.

Sampling Lemons

Here is a simulated sampling distribution of $\bar{x}$ = sample mean lemon weight (oz), based on 500 samples of size *n* = 6 from an approximately normal population with $\mu = 4$ and $\sigma = 0.5$.



When thinking about how you describe a distribution, you should always think about ________________, ________________ and ____________________.

The distribution of sample means is:

- ____
- ____
- ____

Mean and Standard Deviation

Let $\bar{x}$ be the mean of a sample of size *n* selected from a population with mean μ and standard deviation σ. Then,

$$\mu_{\bar{x}} = \mu$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

*assuming the sample size is less than 10% of the population size.

Good News! These are on the formula sheet! Take a minute to find them now!

Name_______________________________

Mean and Standard Deviation
In our example, we selected samples of size *n* = 6 from an approximately normal population with $\mu = 4$ and $\sigma = 0.5$. Therefore: $\mu_{\bar{x}}$ = ________ = __________ $\sigma_{\bar{x}}$ = ________ = __________; assuming 6 is less than _________ of all lemons on the tree.
Shape
Let $\bar{x}$ be the mean of a sample of size *n* selected from a population with mean μ and standard deviation σ. Then, the sampling distribution of $\bar{x}$ will be approximately normal when: • The ________________ distribution is ___________________ normal (as in the lemon example). OR • The ________________ distribution is ________ approximately normal but the ___________ size is ______________________ *e.g., greater than or equal to _________). This is called the _______________________________________ (see Topic 5.3).
What Should We Take Away?
How do we determine the parameters of a sampling distribution of a sample mean? **Mean: $\mu_{\bar{x}} = \mu$** **Standard Deviation: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, assuming the sample size is less than 10% of the population size.** How do we determine if the shape of a sampling distribution of a sample mean is approximately normal? **The ____________________ distribution of $\bar{x}$ will be _________________________ normal when the population distribution is _________________________ normal or the sample size is ___________________ enough ($n \geq 30$).**

Name________________________________

AP Statistics CED 5.7 Daily Video 2 (Skill 4.B)

Sampling Distribution for Sample Means

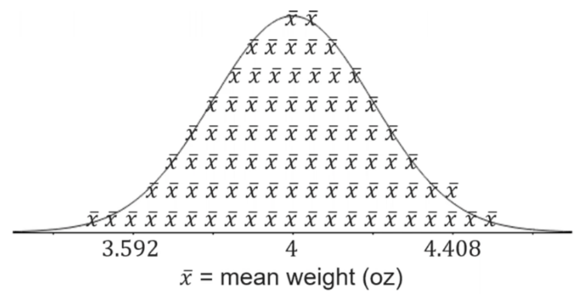
What Will We Learn?

How do we interpret the parameters of sampling distribution of a sample mean?

How do we calculate and interpret probabilities involving a sampling distribution of a sample mean?

Sampling Lemons:

A large tree produces lemons with weights that are approximately normally distributed with a mean of 4 ounces and a standard deviation of 0.5 ounces. Suppose we select a random sample of n = 6 lemons from the tree, weigh each lemon, and calculate the sample mean weight $\bar{x}$ for each sample.



The sampling of $\bar{x}$ is:

- ________________________ (because the _______________ distribution is - approximately normal.
- $\mu_{\bar{x}} = \mu =$ ______
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} =$ _____ = ______

Interpreting the Mean

Interpret the value $\mu_{\bar{x}}$ = 4:

For _________________________ of size n=______ from __________ population, the _____________ mean weights of lemons will have a mean of ______________________.

Interpreting the Standard Deviations

Interpreting the value $\sigma_{\bar{x}}$ = 0.204

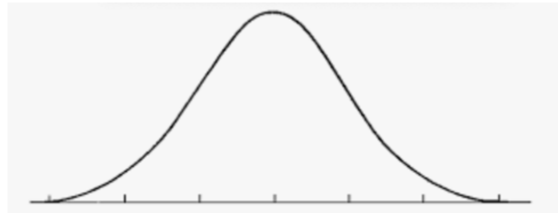
For _________________________ of size n=______ from __________ population, the _____________ mean weights of lemons will _____________ vary by about _____________ from the _______________ mean of ________________. *Note, that knowing how much a _______________________________ typically varies from the ____________ is one of the most important reasons to study _______________ distributions and is the basis of the study of confidence intervals.

Name______________________________

Calculating Probabilities

A large tree produces lemons with weights that are approximately normally distributed with a mean of 4 ounces and a standard deviation of 0.5 ounces. Suppose we select a random sample of n = 6 lemons from the tree. Would it be **unusual** to get a sample mean of $\bar{x}$ = 4.5 or greater in a random sample of size n = 6 from this population?

Step 1: Draw a picture (Sketch from video)



Step 2: Calculate z-score = _____________

Step 3: Calculate $P(\bar{x} \geq 4.5)$ = ____________

Step 4: Interpret the probability – Getting _____________ mean of 4.5 ounces or more happens in only about __________ of ______________ samples of size _____ from this tree. This is _____ unusual.

*Know what values of a statistic count as _______________ is one of the most important reasons to study _________________ distributions.

What Should We Take Away?

How do we interpret the parameters of sampling distribution of a sample mean?

- **In _______________, with ____________**
- **Reference _________________________________ of this size from the population.**
- **For standard deviation, include _______________ or ______________________.**

How do we calculate and interpret probabilities involving a sampling distribution of a sample mean?

- **Use _________________________________ procedures from Topic 5.2 with the parameter of the _____________________ distribution of** $\bar{x}$ (Video 1)
- **Interpretations should be __________________, _________________ and reference ______________________________ of this size from the _________________________.**

Name________________________________

AP Statistics CED 5.8 Daily Video 1 (Skill 3.B)

Sampling Distributions for Difference in Sample Means

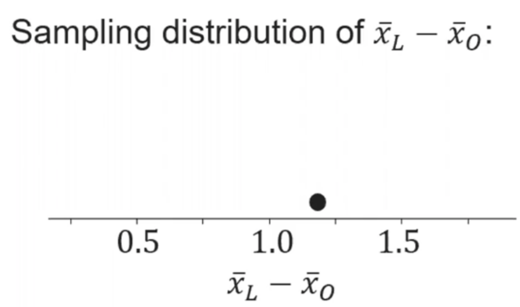
What Will We Learn?

How do we determine the parameters of a sampling distribution of a difference in sample means? How we determine if the shape of a sampling distribution of a difference in sample means is approximately normal?

Sampling Citrus

A large tree produces lemons with weights that are approximately normally distributed with $\mu_L = 4$ ounces and $\sigma_L = .5$ ounces. A different tree produces oranges with weights that are approximately normally distributed with $\mu_O = 3$ ounces and $\sigma_O = .4$ ounces. Suppose we select a random sample of n = 6 fruit from each tree, calculate the mean weight for each sample, and compute the difference in sample means $\bar{x}_L - \bar{x}_O$. (Label and plot possible samples points on graph as you watch video)

One possible sample of lemons: $\bar{x}_L$ = 4.1 ounces
One possible sample of oranges: $\bar{x}_O$ = 2.9 ounces
$\bar{x}_L - \bar{x}_O$ = 1.2 ounces

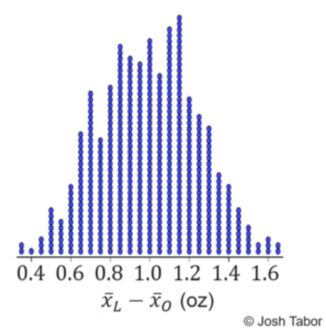


One possible sample of lemons: $\bar{x}_L$ = ____________
One possible sample of oranges: $\bar{x}_O$ = ____________
$\bar{x}_L - \bar{x}_O$ = ____________

One possible sample of lemons: $\bar{x}_L$ = ____________
One possible sample of oranges: $\bar{x}_O$ = ____________
$\bar{x}_L - \bar{x}_O$ = ____________

Of course, these are only three of the MANY possible differences.

Here is a simulated sampling distribution of $\bar{x}_L - \bar{x}_O$ based on 500 random samples of size n_L = 6 from the lemon tree and 500 random samples of size n_O = 6 from the orange tree. How would you describe this distribution?



The distribution of $\bar{x}_L - \bar{x}_O$ is:

- ________________________________
- Centered at ______________________
- Varies from ______________ to about ______________.

Mean and Standard Deviation

Let $\bar{x}_1$ be the sample mean in a random sample of size n_1 selected from population 1 with μ_1 and standard deviation σ_1, and let $\bar{x}_2$ be the sample mean in a random sample of size n_2 selected from population 2 with μ_2 and standard deviation σ_2. Then,

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

*assuming the sample sizes are less than 10% of the population sizes and that the two samples are independent.

STATS MEDIC

Name________________________________

Mean and Standard Deviation

For the lemons, μ_L= 4 ounces and σ_L= .5 ounces.
For the oranges, μ_O= 3 ounces and σ_O= .4 ounces.

Therefore,

$\mu_{\bar{x}_L - \bar{x}_O}$ = __________________ = __________________ = __________________

$\sigma_{\bar{x}_L - \bar{x}_O}$ = __________________ = __________________ = __________________

* Assuming the samples sizes are ______________________ of the population sizes and the two samples are ______________________________.

Mean and Standard Deviation

Good News! These are all on the formula sheet! Take a moment to find them NOW!

Shape

The sampling distribution of $\bar{x}_1 - \bar{x}_2$ will be approximately normal when:

- ________ population distributions are approximately normal (as in the citrus example).

OR

- The populations are _______ approximately normal but both of the sample sizes are _____________________ (e.g., ________________________)

What Should We Take Away?

How do we determine the parameters of a sampling distribution of a difference in sample means?

Mean: $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

SD: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, **assuming the sample sizes are less than 10% of the population sizes and the samples are independent.**

How we determine if the shape of a sampling distribution of a difference in sample means is approximately normal?

The sampling distribution of $\bar{x}_1 - \bar{x}_2$ will be approximately normal when the population Distributions are both approximately normal OR when both sample sizes are large enough (e.g., greater than or equal to 30).

Name_______________________________

AP Statistics CED 5.8 Daily Video 2 (Skill 4.B)

Sampling Distributions for Differences in Sample Means

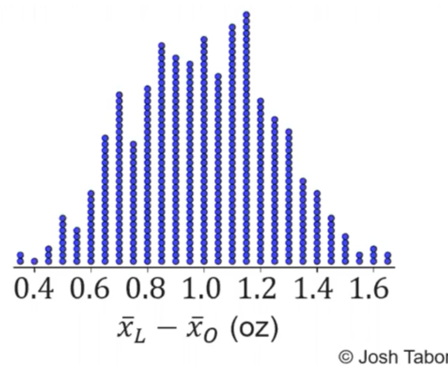
What Will We Learn?

How do we interpret the parameters of a sampling distribution of a difference in sample means?

How do we calculate and interpret probabilities involving a sampling distribution of a difference in sample means?

Sampling Citrus

A large tree produces lemons with weights that are approximately normally distributed with μ_L= 4 ounces and σ_L= .5 ounces. A different tree produces oranges with weights that are approximately normally distributed with μ_O= 3 ounces and σ_O= .4 ounces. Suppose we select a random sample of n = 6 fruit from each tree, calculate the mean weight for each sample, and compute the difference in sample means $\bar{x}_L - \bar{x}_O$.



The distribution of $\bar{x}_L - \bar{x}_O$ *is*:

- ____________________ normal
- $\mu_{\bar{x}_L - \bar{x}_O}$ = ____________ = ____________ =______
- $\sigma_{\bar{x}_L - \bar{x}_O}$ = ____________________________________

Interpreting the Mean

Interpret the value of $\mu_{\bar{x}_L - \bar{x}_O}$ = 1

For ________ random samples of ______________ from the lemon tree and ______________ from the orange tree, the differences (*L – O)* in ________________ mean weight will have a ____________ of ________________.

Interpreting the Standard Deviation

Interpret the value $\sigma_{\bar{x}_L - \bar{x}_O}$ = 0.26.

For _______ random samples of _________________ from the lemon tree and _______________ from the orange tree, the differences (*L – O)* in the _______________ mean weight will _________________ vary by _______________ 0.26 ounces from the _____________ difference in means of 1 ounce.

*Knowing how much a ______________ statistic ________________ varies from the _____________ is one of the most important reasons to study _________________________ distributions!!

Name________________________________

Calculating Probabilities

Suppose we selected a random sample of 6 oranges and 6 lemons. Would it be unusual to get a sample mean weight for the lemons that was more than 1.2 ounces greater than the sample mean weight for the oranges?

We want to find $P(\bar{x}_L - \bar{x}_O > 1.2)$.

Step 1: Draw a picture.

Step 2: Calculate z-score: ___________________

Step 3: Find probability using Table A or Technology: $P(\bar{x}_L - \bar{x}_O > 1.2)$ = ___________________

Interpreting Probabilities

Would it be unusual to get a sample mean weight for the lemons that was more than 1.2 ounces greater than the sample mean weight for the oranges?

Getting a _________________ (L – O) is _______________ mean weights of more that 1.2 ounces happens in _________________________________ of __________ possible samples of _______________ and ________________ from these trees. This is ___________________.
*Knowing what values of a __________________ count as _________________ is one of the most important reasons to study _______________________ distributions.

What Should We Take Away?

How do we interpret the parameters of a sampling distribution of a difference in sample means?

- **In ____________, with _____________**
- **Reference _______ possible samples of ____________ sizes from the populations.**
- **For standard deviation, include __________________ or __________________.**

How do we calculate and interpret probabilities involving a sampling distribution of a difference in sample means?

- **Use _________________ distribution procedures from Topic 5.2 with the parameters of the _______________________ distribution of $\bar{x}_1 - \bar{x}_2$ (Video 1).**
- **Interpretations should be in ___________, with _________ and reference ________ possible samples of these _____________ from the populations.**