Name: ________________________________ AA

STATISTICS UNIT
GMP IV Summer Assignment 2022
SUBMIT BY AUGUST 12, 2022 ON UB LEARNS

The summer assignment will be included in your first quarter grades and includes 13 lessons, 13 homework assignments, 1 activity, 1 test review, 1 quiz, and 1 test.

This is not meant to be done in one sitting. Plan ahead and pace yourself so that you are doing one lesson at a time. You will get much more out of it this way.

You are expected to watch the video for each lesson, complete the notes for each lesson, and complete the assignment for each lesson. You will be submitting, the lesson, assignment, any graphs you make, and any data you collect from the simulation programs on UB learns. Neatly show all work in the space provided as you must show work to get full credit. Clearly print your name in the "Name" section. You must show all work to receive full credit.

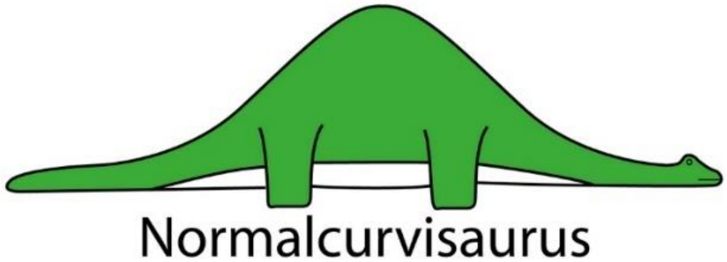
**For lessons 3.5, 13.6, 13.7, and 13.8 you are required to use a simulation program. These can be found at: https://www.emathinstruction.com/courses/common-core-algebra-ii/statistical-simulators/.
You should print a copy of your data and submit it with your work.**

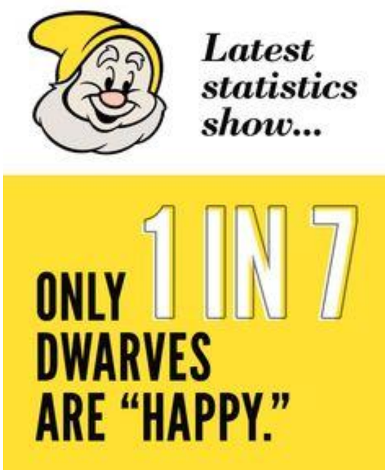
I have included several pages of blank normal curves for lessons, homework problems, quiz/test problems. You are encouraged to use them but not required to use them/

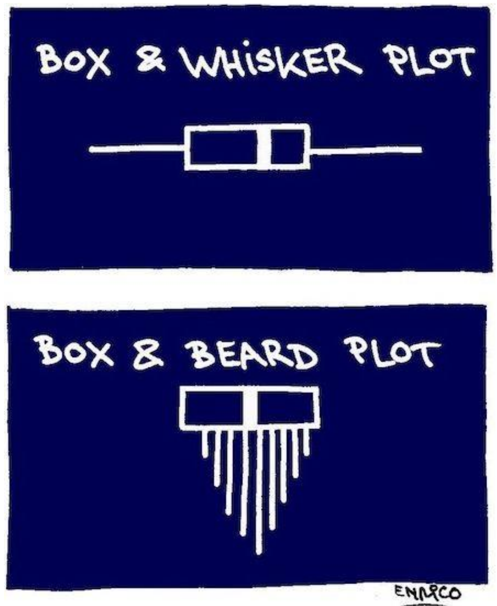
Use the chart below as a checklist and if you have any questions my email address is pshepard@buffalo.edu.

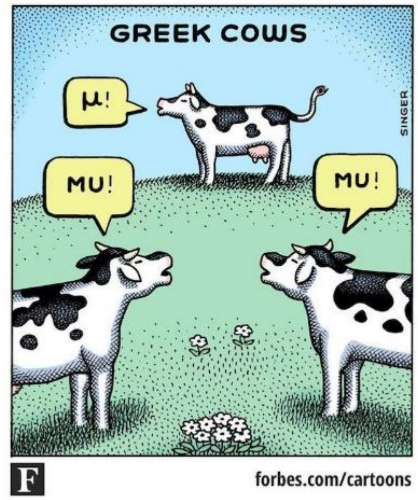
Lesson 13.1 Variability and Sampling		**Video Watched**
		Lesson
		Homework Assignment
Lesson 13.2 Population Parameters		**Video Watched**
		Lesson
		Homework Assignment
Lesson 13.3 The Normal Distributions		**Video Watched**
		Lesson
		Homework Assignment
Lesson 13.4 The Normal Distribution and Z-Scores		**Video Watched**
		Lesson
		Homework Assignment

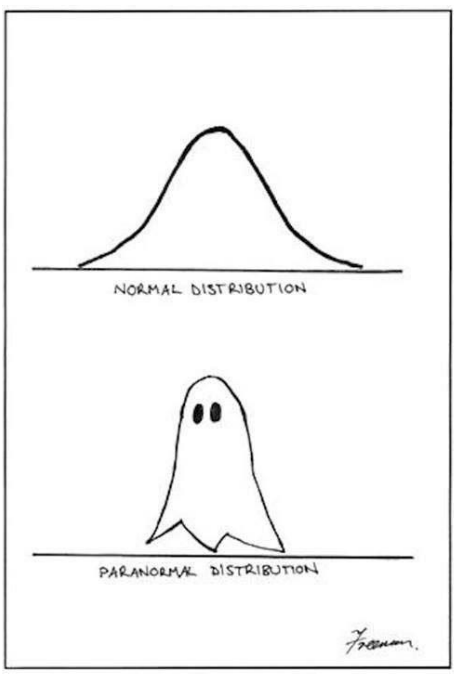
Lesson 13.4.5 Sampling a Population		**Lesson**
		Homework Assignment
Lesson 13.5 Sample Proportions **NORMSAMP program need** **Include a printout of all data**		**Video Watched**
		Lesson
		Homework Assignment
Lesson 13.6 Sample Means **PSIMUL program needed** **Include a printout of all data**		**Video Watched**
		Lesson
		Homework Assignment
Unit 13 Quiz		**Quiz Completed**
Lesson 13.7 The Difference in Samples Means **MEANCOMP program needed** **Include a printout of all data**		**Video Watched**
		Lesson
		Homework Assignment
Lesson 13.8 The Distribution of Sample Means **NORMSAMP program need** **Include a printout of all data**		**Video Watched**
		Lesson
		Homework Assignment
Lesson 13.9 The Distribution of Sample Proportions		**Video Watched**
		Lesson
		Homework Assignment
Unit 13: Statistical Simulation Packet		**Activity Completed**
Lesson 13.10 The Distribution of Sample Proportions		**Video Watched**
		Lesson
		Homework Assignment
Lesson 13.11: Linear Regression and Lines of Best Fit		**Video Watched**
		Lesson
		Homework Assignment
Lesson 13.12 Other Types of Regression		**Video Watched**
		Lesson
		Homework Assignment
Unit 13 Review		**Completed**
Unit 13 Test		**Completed**





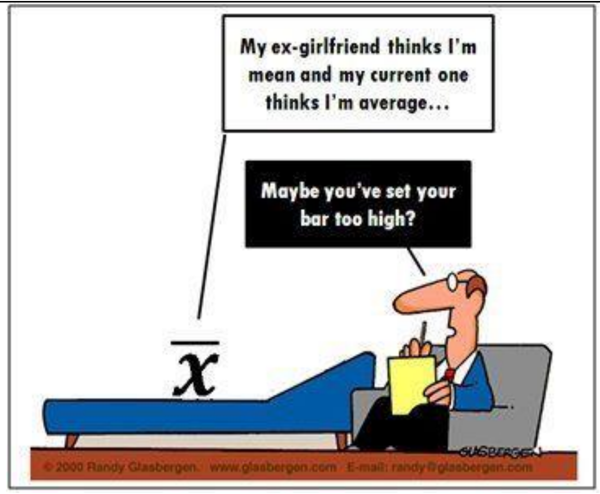


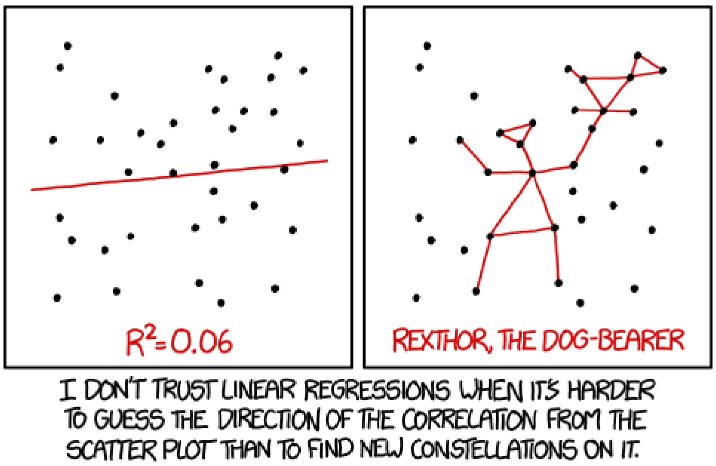


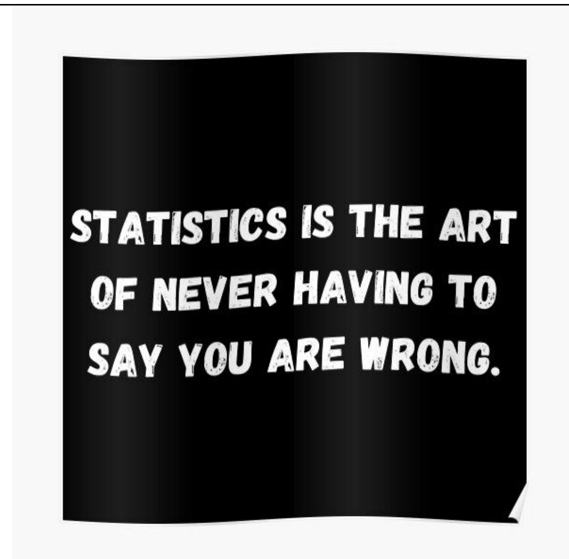


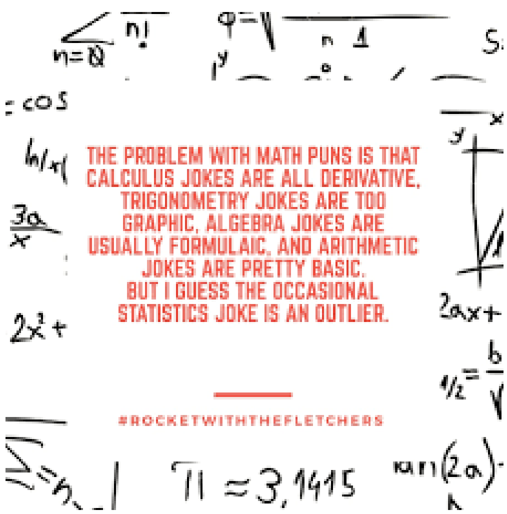
"SHOULD WE SCARE THE OPPOSITION BY ANNOUNCING OUR MEAN HEIGHT OR LULL THEM BY ANNOUNCING OUR MEDIAN HEIGHT?"

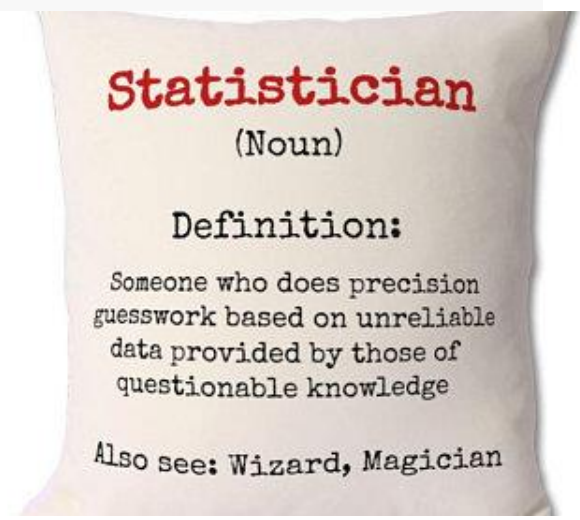

DEPARTMENT OF PROBABILITY
VERY LIKELY TO BE BACK IN AN HOUR
CARTOONSTOCK .com
Search ID: msen295
Seddon..


My ex-girlfriend thinks I'm mean and my current one thinks I'm average...
Maybe you've set your bar too high?
$\bar{x}$
GLASBERGEN
© 2000 Randy Glasbergen. www.glasbergen.com E-mail: randy@glasbergen.com

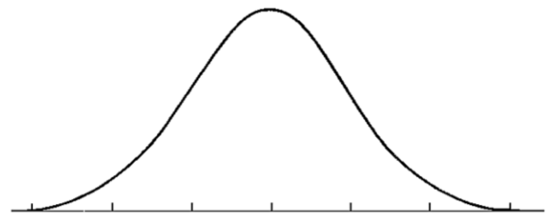

$R^2=0.06$
REXTHOR, THE DOG-BEARER
I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

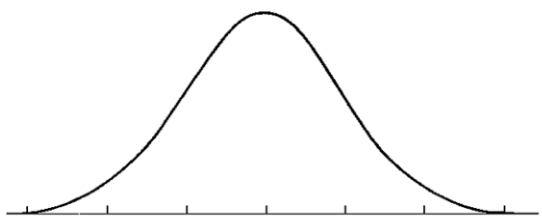

STATISTICS IS THE ART OF NEVER HAVING TO SAY YOU ARE WRONG.

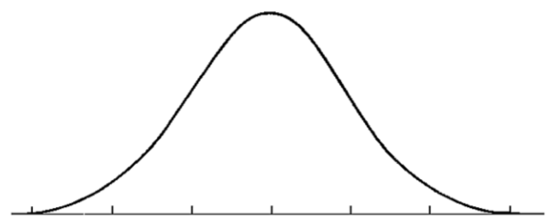

THE PROBLEM WITH MATH PUNS IS THAT CALCULUS JOKES ARE ALL DERIVATIVE, TRIGONOMETRY JOKES ARE TOO GRAPHIC, ALGEBRA JOKES ARE USUALLY FORMULAIC, AND ARITHMETIC JOKES ARE PRETTY BASIC.
BUT I GUESS THE OCCASIONAL STATISTICS JOKE IS AN OUTLIER.
#ROCKETWITHTHEFLETCHERS

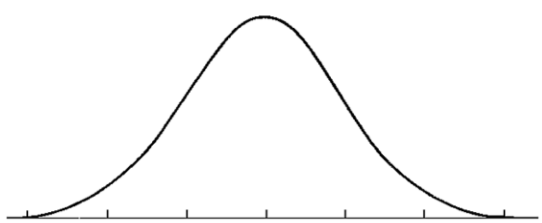

Statistician
(Noun)
Definition:
Someone who does precision guesswork based on unreliable data provided by those of questionable knowledge
Also see: Wizard, Magician

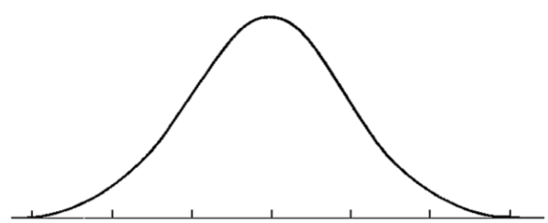
Name: ________________________________

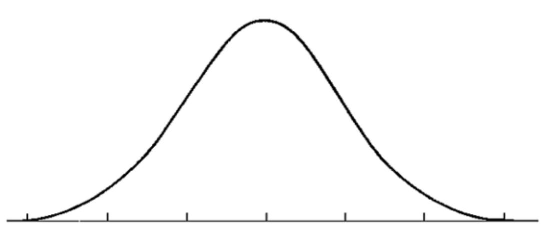


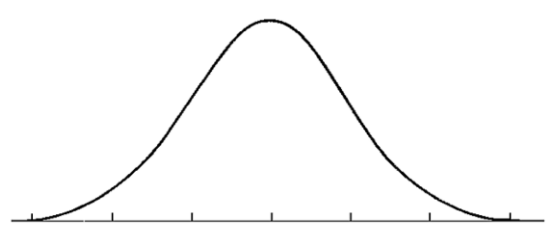




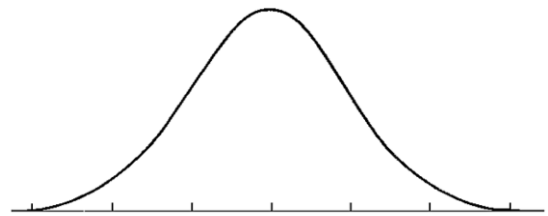


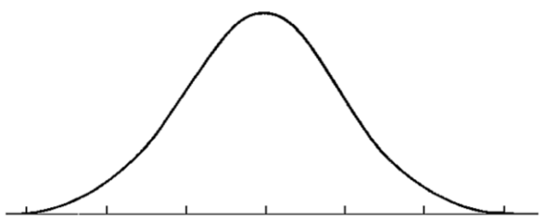


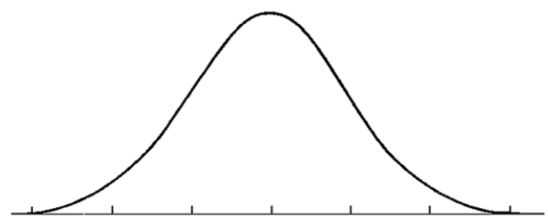


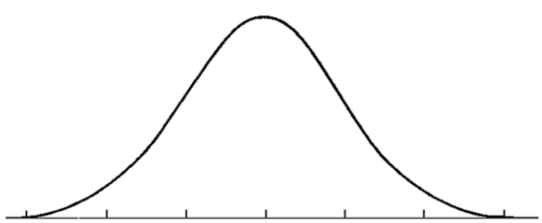


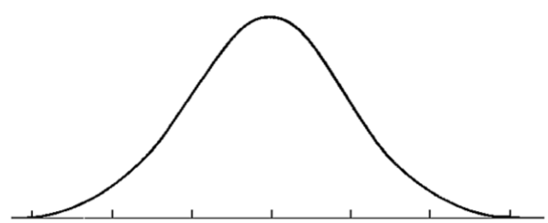
Name: ________________________________

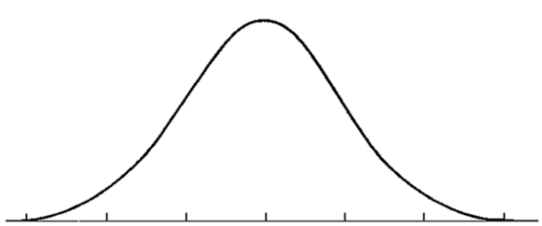


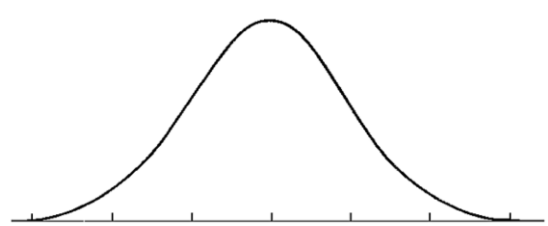




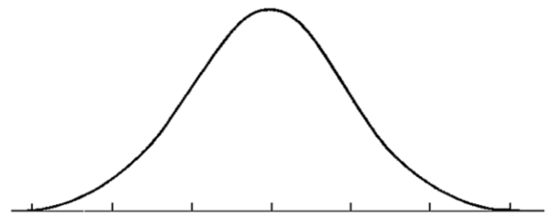
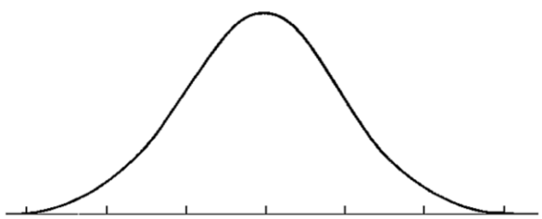
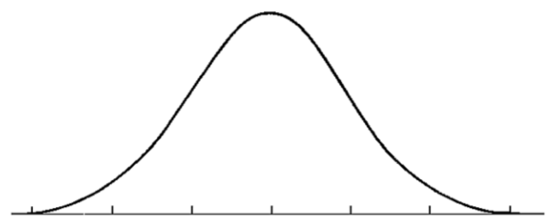
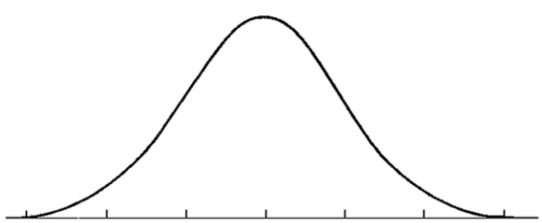
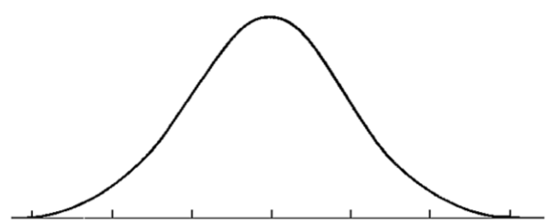
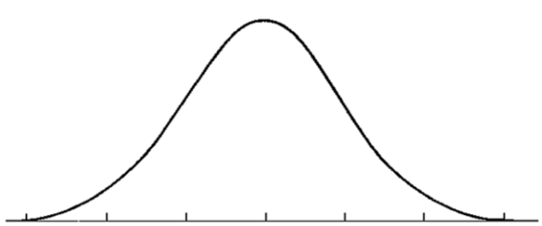
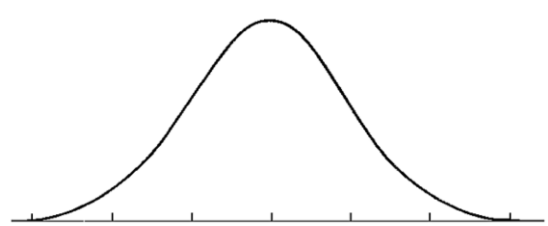




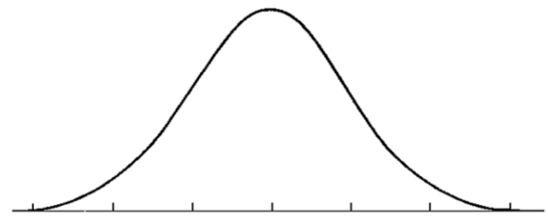


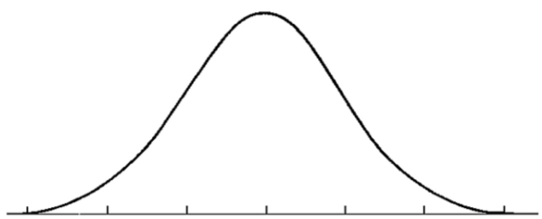


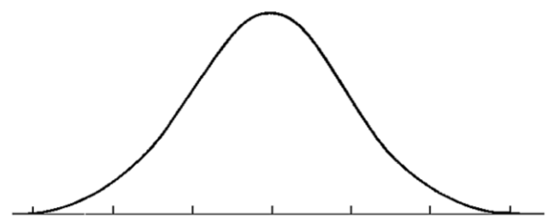
Name: ________________________________

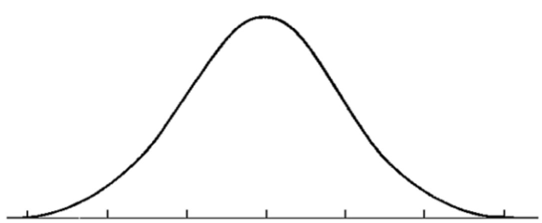









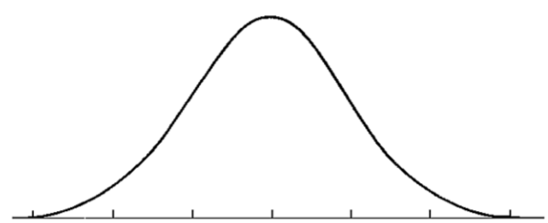
Name: ______________________________

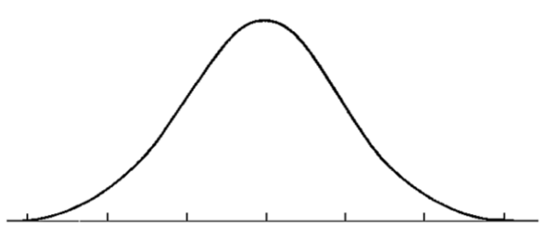


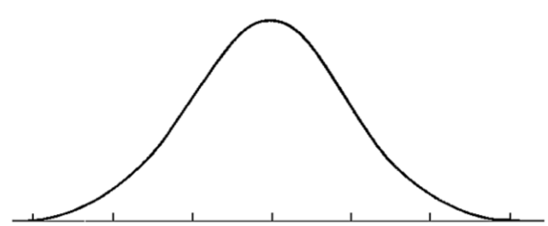




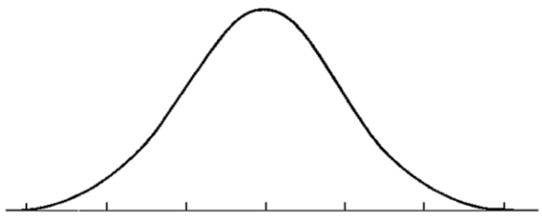
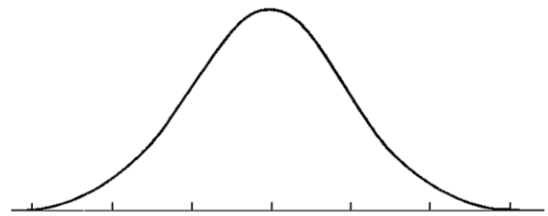
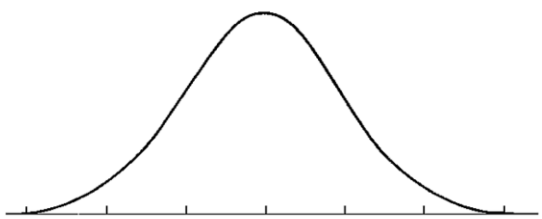
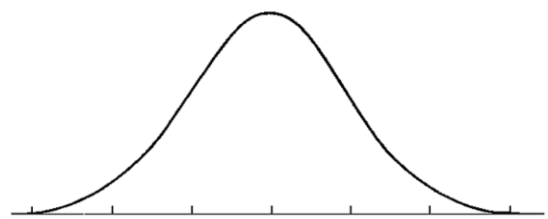
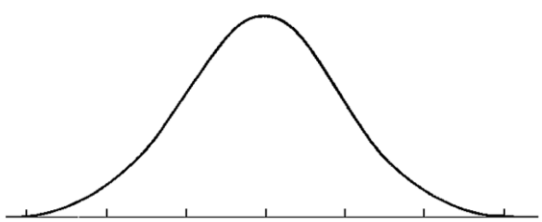
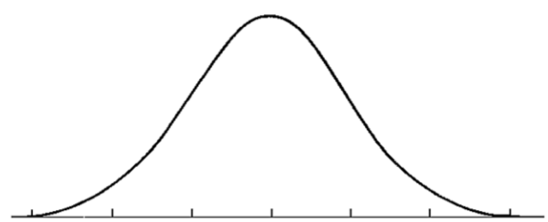
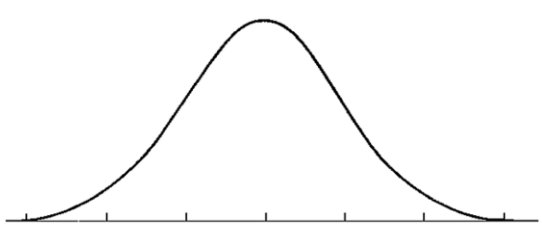
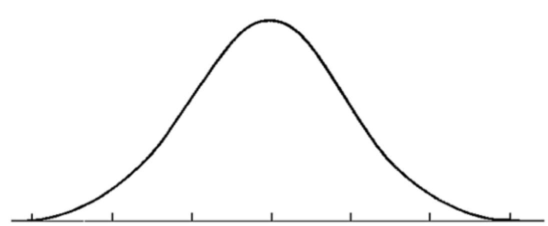








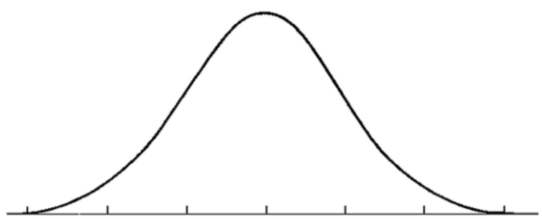
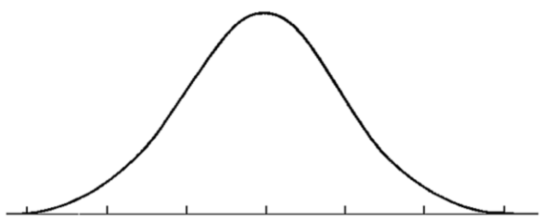
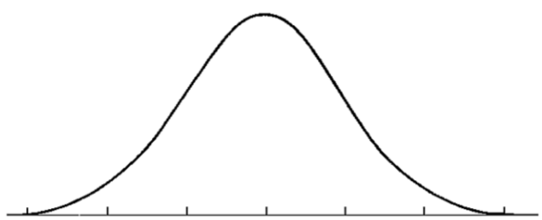
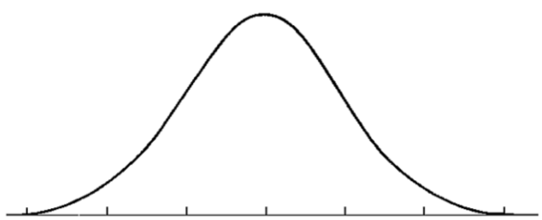
Name: ________________________________

Name: ________________________________





Name: ___________________________________ Date: _________________

VARIABILITY AND SAMPLING
COMMON CORE ALGEBRA II

Data is everywhere. It's in our newspapers, it's in our science classes, it shows up in economics, medicine and anywhere else that **variability** occurs. Variability is simply the property of **outcomes** being different. The tools of statistics are designed to **explain this variability**.

There are many types of variability. It is good to understand these sources in order to minimize the ones that we are not studying.

***Exercise* #1:** The following types of variability can change uniformity of a data set. For each give an example from any field.

(a) **Observational or Measurement Variability**: Variability that is introduced due to either our measuring instruments not being precise enough or differences in how two different people read the measurement.

(b) **Natural Variability or Inter-Individual Variability**: Variability that accounts for the fact that members of a population are simply different.

(c) **Induced Variability**: This type of variability is in marked contrast to natural. It occurs because we have assigned our population or sample to two or more **treatment** groups and then observe the variability between the groups.

(d) **Sample Variability**: This is the type of variability that occurs when we take multiple **samples** from a **population** randomly. These samples will be different due to the randomness of the sampling process.

Remember, through all of our work in this unit, we are really trying to explain the variability of data within either a population or a sample and then using this to determine if the variability can be attributed to one of the factors above to the exclusion of the others.

There are many different situations in which we collect data. They have important differences and all of them depend on **randomization** in one way or another.

***Exercise* #2:** The three major types of ways to collect data are described below. Give an example of each and explain how **randomization** is part of each method. Randomization is used primarily to eliminate variability caused by some type of **bias**.

(a) **Surveys**: Collections of data from a population where variability is not induced by treatments but by the sample itself (sampling variability).

(b) **Observational Studies**: Collections of data from a population where assignment of individuals from the population into **treatment groups** is **not** under the control of those performing the study.

(c) **Experimental Studies**: In experimental studies individuals are assigned randomly to treatment groups in order to determine the effect of the treatment on the variability of the data. In these cases, the assignment, although random, is under the control of those performing the study.

Random sampling is critical for being able to minimize variability due to **sampling bias**. Random sampling can be done using a variety of different techniques. Simple random sampling can be accomplished using a random number table.

***Exercise* #3:** A list of 10 people's heights, in inches, is shown below.

Person #	1	2	3	4	5	6	7	8	9	10
Height	70	68	60	75	65	69	58	62	66	63

(a) Randomly select five heights from this list by using the random number table that goes with this lesson. Choose a random spot in the table and move down the column. Select the first digit of each number. If you get a repeat, eliminate and keep going. If you get a 0, use this as the 10.

(b) Calculate the **sample mean** to the nearest tenth. Compare to others in the class. What type of variability is being introduced through this process?

Name: ___________________________________ Date: ___________________

VARIABILITY AND SAMPLING
COMMON CORE ALGEBRA II HOMEWORK

APPLICATIONS

1. Scientists randomly select ten groups from a population of men over 50 years old. They calculate the mean weights of each of these groups. The variability between these means can be best attributed to

   (1) measurement variability (3) induced variability

   (2) natural variability (4) sampling variability

   ____

2. Max and Daniel are measuring the amount of time it takes for a ball to roll down a ramp at different heights. For each trial, both Max and Daniel take turns rolling the ball and working the stop watch. They do this in order to quantify which of the following sources of variability?

   (1) measurement variability (3) induced variability

   (2) natural variability (4) sampling variability

   ____

3. Which of the following scenarios would be an attempt to quantify induced variability?

   (1) a phone survey of political preferences during election season
   (2) multiple random samples of products from an assembly line to check for defects
   (3) random assignment of people to a control group and a group taking a drug to lower cholesterol
   (4) recording the variability in the measurement of a soil sample's weight by the same machine

   ____

4. Which of the following research questions would involve collecting data through a survey?

   (1) Watching people exit a grocery store to see the percent who use reusable bags.
   (2) Assigning people to two groups to see the effect of a particular amount of sleep.
   (3) Calling people on the telephone to see if they will be voting in the upcoming election.
   (4) Dropping salt cubes into two different liquids to determine which dissolves faster.

   ____

5. In which of the following cases would an observational study be necessary as compared to an experimental study?

   (1) The study of how increased nutrient levels affect plant growth.
   (2) The study of how educational levels affect median household income.
   (3) The study of how a vaccine affects the percent of mice that get a particular disease.
   (4) The study of how noise level affects the sleep patterns of volunteers in a sleep study.

   ____

6. In an experimental study, a lab wanted to divide volunteers into two groups to determine the effect of a particular phone app to help make people more punctual (on time). The 50 volunteers in the study will be assigned to either a group of 25 who use the app for a week or a group of 25 who do not use the app. The participants were asked to come to a lab to receive the app (or not) at 10:00 am on a Monday. Answer the following questions:

   (a) Why would those performing the study *not* want to assign the participants in the two treatments (groups) based on who showed up to the study session first?

   (b) Propose a way to use a random number table to generate a simple random selection that eliminates the bias that you discussed in part (a).

7. Two groups of subjects were divided in an experimental study. One group was given a drug to help speed up their metabolism and result in weight loss. The other group was given a **placebo** (a pill that looks identical to the one given to the other group, but without the weight loss drug). After a month of the experiment, the weight loss of each individual in each of the two groups were measured. In general, people in the group who took the metabolism drug did lose more weight, although there were differences in the amount each lost. There are two main types of variability occurring in this study. Describe each type below in the context of this study.

   Induced Variability

   Natural Variability

REASONING

8. If you were trying to conduct a survey of political preferences for likely voters in an upcoming election and decided to dial 1,000 randomly generated land-line phone numbers (not cell), why might this still introduce **bias** into the sampling?

Random Number Table

89679	74452	58378	56038	05793
68479	31125	30744	92830	81733
54958	34875	26881	95459	05001
09728	86396	44698	00445	54666
49645	05086	43332	07908	10593
97742	58396	05140	74052	42483
60394	75922	71275	85120	29034
36606	75808	63047	96796	99834
24656	44208	95016	79816	14185
99387	64057	29448	78761	90544
85213	94939	36368	06737	30994
01727	01497	49402	88141	58513
57535	40645	17498	85894	03705
29613	07446	68202	19465	79334
74042	64704	75418	80166	50113
05561	96960	41774	27701	26791
13709	71189	29285	16286	67827
57752	35321	45784	58222	99383
87272	68090	81526	13161	80735
28664	27875	78093	30888	92618
85995	57330	24519	17238	34929
19402	86361	97351	89230	84306
25335	15291	13878	89663	82143
19631	14030	58249	22092	10967
29731	65359	83185	55700	09254
12342	51338	50542	47077	99987
81333	34849	35289	04468	60304
14825	35419	03873	09164	25581
47865	82527	72916	69732	62153
46246	21019	29652	36296	80016
88454	58304	64450	39653	54792
18412	23667	49507	75752	66366
08044	32980	67699	00755	82771
03017	69707	56600	37524	58097
62259	24785	87969	53877	77589
25294	83064	13116	40659	90535
76449	44295	97098	18216	46682
73964	06143	86782	34176	21466
63960	70532	19083	87598	70803
89628	99681	41047	35674	88642

Name: ___________________________________ Date: _________________

POPULATION PARAMETERS
COMMON CORE ALGEBRA II

When we conduct a study, the complete set of all subjects that share a common characteristic that is being studied is known as the **population**. All populations have **natural or inter-individual variability**. Most of the time, the entire population is not measured, but a sample is taken to infer characteristics of a population. Still, all populations in theory have **population parameters** that describe the population, such as its mean, standard deviation, and interquartile range.

***Exercise* #1:** 18 students in Mr. Weiler's Advanced Calculus class took a quiz with the following results in ascending order.

56, 68, 72, 72, 75, 78, 80, 84, 84, 85, 88, 88, 90, 93, 95, 99, 100, 100

(a) Use your calculator to determine the mean, the median, and the quartiles for this data set. Then, construct a simple box-and-whiskers (box plot) for this data set.

(b) What is the interquartile range of this data set? In theory, what percent of the data set should lie between the first and third quartiles? Is that true for this data set?

(c) What is the population standard deviation for this data set to the nearest tenth? How do you interpret the standard deviation?

(d) What percent of the scores were within one standard deviation of the mean? Within two standard deviations of the mean? Round your percents to the nearest percent and show your work.

Within One Standard Deviation of the Mean

Within Two Standard Deviations of the Mean

Sometimes data is grouped in a frequency chart. We still should be able to calculate the basic population parameters when the information is given in this form.

***Exercise* #2:** A small company has salaries for their 50 employees as given in the table below

Salary (x_i)	Frequency (f_i)
25,000	5
32,000	21
45,000	14
58,000	7
75,000	2
120,000	1

(a) Find the mean and standard deviations of the salary range.

(b) What is the median of this data set? Why is the median considerably lower than the mean in this data set?

(c) Does more or less than 50% of the data set fall within one standard deviation of the mean? Show the analysis that leads to your answer.

Although we have often concentrated on experimental studies where data is collected and means are found, many times we use statistics to represent results of a survey where we are interested in what **proportion** of a **population** share a certain characteristic. These proportions are mostly expressed as decimals, but sometimes are represented by fractions or percents.

***Exercise* #3:** A questionnaire went home to all juniors concerning their ability to bring and use mobile devices at school. The questionnaires constituted a **census** since all of the juniors were surveyed. Of the 742 juniors, 564 of them reported having web-enabled mobile devices. What was the population proportion for web-enabled devices? Express your answer as a decimal and as a percent.

***Exercise* #4:** The proportion of eggs that get cracked in a local egg handling facility is 0.023. If 2,500 dozen eggs are packaged in the factory per day, what should we expect to be the number of eggs cracked per day?

(1) 350 (3) 230

(2) 450 (4) 690

Name: ____________________________________ Date: __________________

POPULATION PARAMETERS
COMMON CORE ALGEBRA II HOMEWORK

FLUENCY

1. Which of the following formulas, written in summation notation, would represent the mean of the data set $\{x_1, x_2, ..., x_n\}$? Explain your choice.

(1) $\sum_{i=1}^{n} x_i$ (3) $n\sum_{i=1}^{n} x_i$

(2) $\frac{1}{n}\sum_{i=1}^{n} x_i^2$ (4) $\frac{1}{n}\sum_{i=1}^{n} x_i$ __________

2. The standard deviation of a population characteristics measures

(1) The difference between the maximum and minimum values.
(2) The difference between the third quartile and first quartile values.
(3) The average distance a data value is away from the mean.
(4) The average distance a data value is away from the median. __________

3. The interquartile range of the data set $\{4, 7, 10, 13, 18, 22, 30\}$ is

(1) 15 (3) 7

(2) 18 (4) 10 __________

APPLICATIONS

4. If 348 freshmen out of 622 have cell phones, then the population proportion, p, for freshmen cell phone ownership is

(1) 0.56 (3) 0.72

(2) 0.35 (4) 0.44 __________

5. If a population has 824 subjects, then about how many would have characteristics in the upper quartile?

(1) 412 (3) 368

(2) 280 (4) 206 __________

6. A school is tracking its freshmen attendance for the first marking period. Shown below is a table summarizing their findings for the 284 members of the freshmen class.

Days Absent (x_i)	Number of Students (f_i)
0	158
1	64
2	18
3	22
4	4
5	7
6	8
9	2
13	1

(a) Find the mean and median number of days absent. Round your mean to the nearest tenth.

(b) What is the population standard deviation for this data set? Round to the nearest tenth.

(c) What proportion of the population that has an absenteeism greater than 4 days?

7. The heights of the 15 players on the Arlington boys' varsity basketball team are given below in inches.

66, 67, 68, 68, 70, 72, 72, 73, 74, 75, 75, 75, 76, 77, 79

(a) Find the mean and standard deviation of this data set. Use the population standard deviation. Round both to the nearest *tenth*.

(b) Determine the proportion of the population that falls within one standard deviation and within two standard deviations of the mean. State your values in decimal form.

One standard deviation from the mean:

Two standard deviations from the mean:

(c) Use the random number table for this lesson to pick a random sample of five players from this list. Do this by picking a random two digit column along the page. Scan down the column until you have picked 5 random integers that fall from 1 to 15. Write down your sample and calculate its mean.

Name: ______________________________ Date: ______________

THE NORMAL DISTRIBUTION
COMMON CORE ALGEBRA II

Many populations have a distribution that can be well described with what is known as **The Normal Distribution** or the **Bell Curve**. This curve, as seen in the accompanying handout to this lesson, shows the percent or proportion of a normally distributed data set that lies certain amounts from the mean.

***Exercise* #1:** For a population that is normally distributed, find the percentage of the population that lies

(a) within one standard deviation of the mean.

(b) within two standard deviations of the mean.

(c) more than three standard deviations away from the mean.

(d) between one and two standard deviations above the mean.

As can be easily seen from *Exercise* #1, the majority of any normally distributed population will lie within one standard deviation of its mean and the vast majority will lie within two standard deviations. A whole variety of problems can be solved if we know that a population is normally distributed.

***Exercise* #2:** At Arlington High School, 424 juniors recently took the SAT exam. On the math portion of the exam, the mean score was 540 with a standard deviation of 80. If the scores on the exam were normally distributed, answer the following questions.

(a) What percent of the math scores fell between 500 and 660?

(b) How many scores fell between 500 and 660? Round your answer to the nearest whole number.

(c) If Evin scored a 740 on her math exam, what percent of the students who took the exam did better than her?

(d) Approximately how many students did better than Evin?

***Exercise* #3:** The heights of 16 year old teenage boys are normally distributed with a mean of 66 inches and a standard deviation of 3. If Jabari is 72 inches tall, which of the following is closest to his height's percentile rank?

(1) 85^{th} (3) 98^{th}

(2) 67^{th} (4) 93^{rd} ________

***Exercise* #4:** The amount of soda in a standard can is normally distributed with a mean of 12 ounces and a standard deviation of 0.6 ounces. If 250 soda cans were pulled by a company to check volume, how many would be expected to have less than 11.1 ounces in them?

(1) 17 (3) 28

(2) 23 (4) 11 ________

***Exercise* #5:** Biologists are studying the weights of Red King Crabs in the Alaskan waters. They sample 16 crabs and compiled their weights, in pounds, as shown below.

9.8, 10.1, 11.1, 12.4, 11.8, 13.2, 12.8, 12.5, 13.7, 11.6, 13.4, 12.3, 12.6, 14.8, 14.2 15.1

(a) Determine the mean and sample standard deviation for this sample of crabs. Round both statistical measures to the nearest *tenth* of a pound.

(b) Why does this sample indicate that the population would be well modeled using a normal distribution? Explain. Hint – Use your calculator to sort this data in ascending order.

(c) Assuming your mean and standard deviation from part (a) apply to a normally distributed population of crabs caught in Alaska, what percent will fall between 9.6 pounds and 15.6 pounds?

(d) If fishermen must throw back any crab caught below 10.4 pounds, approximately what percent of the crabs caught will need to be thrown back if the weights are normally distributed?

Name: ______________________________________ Date: _________________

THE NORMAL DISTRIBUTION
COMMON CORE ALGEBRA II HOMEWORK

FLUENCY

1. A variable is normally distributed with a mean of 16 and a standard deviation of 6. Find the percent of the data set that:

(a) is greater than 16 (b) falls between 10 and 22 (c) is greater than 28

(d) is less than 1 (e) falls between 4 and 19 (f) falls between 22 and 31

APPLICATIONS

2. The weights of Siamese cats are normally distributed with a mean of 6.4 pounds and a standard deviation of 0.8 pounds. If a breeder of Siamese cats has 128 in his care, how many can he expect to have weights between 5.2 and 7.6 pounds?

(1) 106 (3) 98

(2) 49 (4) 111 __________

3. If one quart bottles of apple juice have weights that are normally distributed with a mean of 64 ounces and a standard deviation of 3 ounces, what percent of bottles would be expected to have less than 58 ounces?

(1) 6.7% (3) 0.6%

(2) 15.0% (4) 2.3% __________

4. Historically daily high temperatures in July in Red Hook, New York, are normally distributed with a mean of 84°F and a standard deviation of 4°F. How many of the 31 days of July can a person expect to have temperatures above 90°F?

(1) 6 (3) 9

(2) 2 (4) 4 __________

5. The weights of four year old boys are normally distributed with a mean of 38 pounds and a standard deviation of 4 pounds. Which of the following weights could represent the 90^{th} percentile for the weight of a four year old?

(1) 47 pounds (3) 43 pounds

(2) 45 pounds (4) 41 pounds

____

6. The lengths of songs on the radio are normally distributed with a mean length of 210 seconds. If 38.2% of all songs have lengths between 194 and 226 seconds, then the standard deviation of this distribution is

(1) 16 seconds (3) 8 seconds

(2) 32 seconds (4) 64 seconds

____

7. The heights of professional basketball players are normally distributed with a standard deviation of 5 inches. If only 2.3% of all pro basketball players have heights above 7 foot 5 inches, then which of the following is the mean height of pro basketball players?

(1) 6 feet 5 inches (3) 6 feet 10 inches

(2) 6 feet 2 inches (4) 6 feet 7 inches

____

8. On a recent statewide math test, the raw score average was 56 points with a standard deviation of 18. If the scores were normally distributed and 24,000 students took the test, answer the following questions.

(a) What percent of students scored below a 38 on the test?

(b) How many students scored less than a 38?

(c) If the state would like to scale the test so that a 90% would correspond to a raw score that is one and a half standard deviations above the mean, what raw score is needed for a 90%?

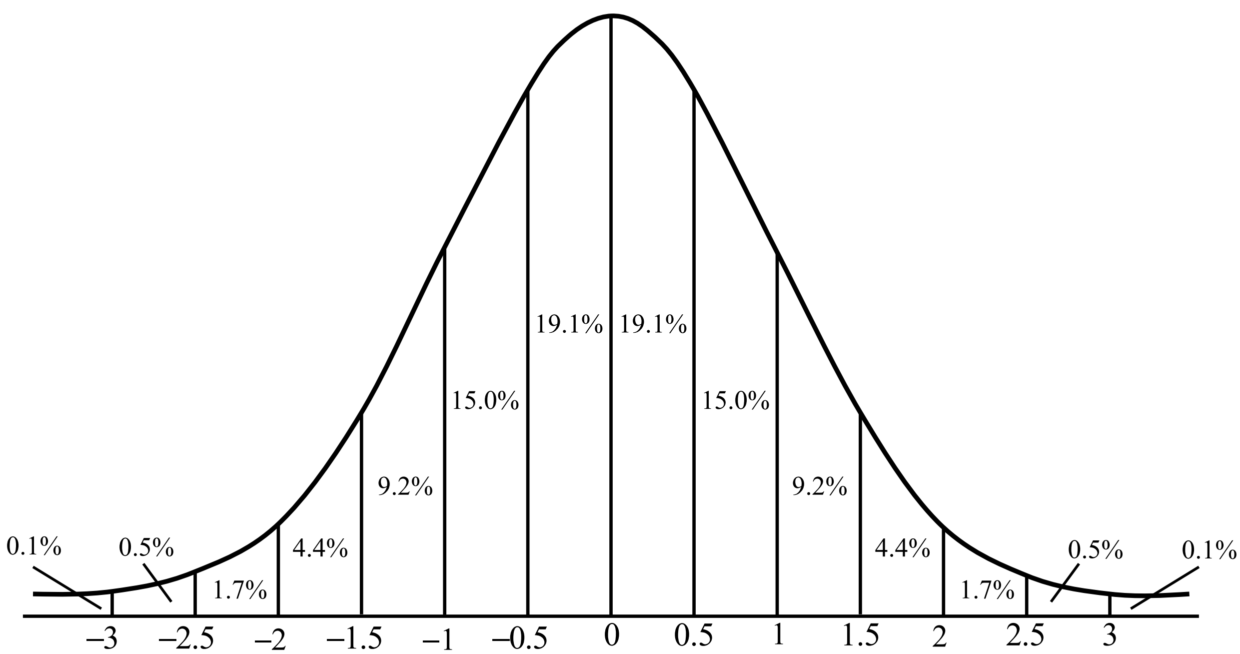
(d) How many of the 24,000 students receive a scaled score greater than a 90%?

(e) The state would like no more than 550 of the 24,000 students to fail the exam. What percent of the total does the 550 represent? Round to the nearest tenth of a percent.

(f) What should the raw passing score be set at so that no more than the 550 students fail?

THE NORMAL DISTRIBUTION

BASED ON STANDARD DEVIATION



Name: ____________________________________ Date: __________________

THE NORMAL DISTRIBUTION AND Z-SCORES
COMMON CORE ALGEBRA II

The normal distribution can be used in increments other than half-standard deviations. In fact, we can use either our calculators or tables to determine probabilities (or proportions) for almost any data value within a normally distributed population, as long as we know the population mean, μ, and the population standard deviation, σ. But, first, we will introduce a concept known as a data value's z-score.

THE Z-SCORE OF A DATA VALUE

For a data point x_i, its z-score is calculated by: $z=\dfrac{x_i-\mu}{\sigma}$. It calculates how far from the mean, in terms of standard deviations, a data point lies. It can be positive if the data point lies above the mean or negative if the data point lies below the mean.

***Exercise* #1:** Boy's heights in seventh grade are normally distributed with a mean height of 62 inches and a standard deviation of 3.2 inches. Find z-scores, rounded to the nearest hundredth, for each of the following heights. Show the calculation that leads to your answer.

(a) $x_i = 66$ inches

(b) $x_i = 57$ inches

(c) $x_i = 70$ inches

Z-scores give us a way to compare how far a data point is away from its mean in terms of standard deviations. We should be able to compute a z-score for a data value and go in the opposite direction.

***Exercise* #2:** Jeremiah took a standardized test where the mean score was a 560 and the standard deviation was 45. If Jeremiah's score resulted in a z-value of 1.84, then what was Jeremiah's score to the nearest whole number?

With z-scores, we can then determine the probability that a subject picked from a normally distributed population would have a characteristic in a certain range. Z-score tables come in many different varieties. The one that comes with this lesson shows only the right hand side, so symmetry will have to be used to determine probabilities.

***Exercise* #3:** The lengths of full grown sockeye salmon are normally distributed with a mean of 29.2 inches and a standard deviation of 2.4 inches.

(a) Find z-scores for sockeye salmon whose lengths are 25 inches to 32 inches. Round to the nearest hundredth.

(b) Use the z-score table to determine the proportion of the sockeye salmon population, to the nearest percent, that lies between 25 inches and 32 inches. Illustrate your work graphically.

***Exercise* #4:** If the scores on a standardized test are normally distributed with a mean of 560 and a standard deviation of 75. Answer the following questions by using z-scores and the normal distribution table.

(a) Find the probability that a test picked at random would have a score larger than 720. Round to the nearest tenth of a percent.

(b) Find the probability that a completed test picked at random would have a score less than 500. Round to the nearest tenth of a percent.

(c) Find the probability that a completed test picked at random would have a score between 500 and 600.

(d) Find the probability that a completed test picked at random would have a score between 600 and 700.

This process is sometimes used to determine a particular data point's **percentile**, which is the **percent of the population less than the data point**.

***Exercise* #5:** The average weight of full grown beef cows is 1470 pounds with a standard deviation of 230 pounds. If the weights are normally distributed, what is the percentile rank of a cow that weighs 1,750 pounds?

(1) 89^{th}

(3) 49^{th}

(2) 76^{th}

(4) 35^{th}

Your graphing calculator can also find these proportions or percent values. Each calculator's inputs and language will be slightly different, although many will do much of the work for you, even allowing you to **not think about the z-scores**.

***Exercise* #6:** Given that the volume of soda in a 12 ounce bottle from a factory varies normally with a mean of 12.2 ounces and a standard deviation of 0.6 ounces, use your calculator to determine the probability that a bottle chosen at random would have a volume:

(a) Greater than 13 ounces.

(b) Less than 11 ounces

(c) Between 11.5 and 12.5 ounces

Name: ____________________________________ Date: __________________

THE NORMAL DISTRIBUTION AND Z-SCORES
COMMON CORE ALGEBRA II HOMEWORK

FLUENCY

1. A population has a mean of $\mu = 24.8$ and a standard deviation of $\sigma = 4.2$. For each of the following data values, calculate the z-value to the nearest hundredth. You do *not* need to read the Normal table.

(a) $x_i = 30$ (b) $x_i = 35$ (c) $x_i = 19$

(d) $x_i = 15.4$ (e) $x_i = 24.8$ (f) $x_i = 33.2$

2. A population has a mean of 102.8 and a standard deviation of 15.4. If a data point has a z-value of 1.87 then which of the following is the value of the data point?

(1) 28.8 (3) 131.6

(2) 86.7 (4) 152.3

____

APPLICATIONS

Get practice with both the Normal Distribution Table and your calculator when doing the following problems.

3. A recent study found that the mean amount spent by individuals on a music service website was normally distributed with a mean of $384 with a standard deviation of $48. Which of the following gives the proportion of the individuals that spend more than $400?

(1) 0.43 (3) 0.12

(2) 0.74 (4) 0.37

____

4. The hold time experienced by people calling a government agency was found to be normally distributed with a mean of 12.4 minutes and a standard deviation of 4.3 minutes. Which percent below represents the percent of calls answered in less than 5 minutes?

(1) 4.3% (3) 6.8%

(2) 5.3% (4) 12.9%

____

5. The national average price per gallon for gasoline is normally distributed with a mean (currently) of \$2.34 per gallon with a standard deviation of \$0.26 per gallon. Which of the following represents the proportion of the gas prices that lie between \$2.00 and \$3.00?

(1) 56% (3) 84%

(2) 72% (4) 90%

______

6. If the average teacher salary in the United States is \$45,753 and salaries are normally distributed with a standard deviation of \$7890, would a salary of \$40,000 per year be in the lowest quintile of teacher salary? (Do a quick Internet search on the term quintile if you don't know what it means).

7. The average rent for a one bedroom apartment (in the Winter of 2015) in New York City is a whopping \$2801 per month with a standard deviation of \$920.

(a) If rents are normally distributed, what percent of the apartments will be less than \$2,500 per month?

(b) If rents are normally distributed, how realistic is it to believe you will be able to rent a one-bedroom in New York City for less than \$1,500 per month? Justify your answer.

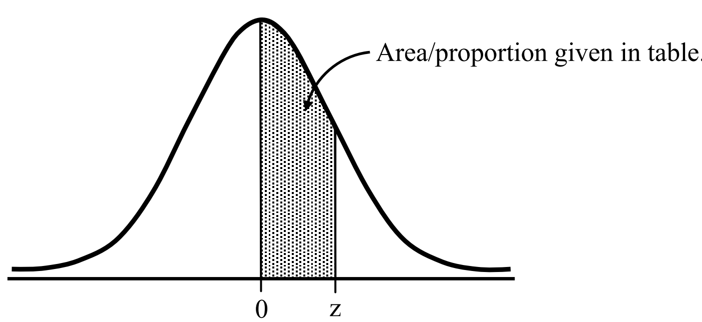
(c) A one-bedroom on the Upper East Side with a doorman and views of Central Park was listed at \$5,000 per month. How rare is this? Assume the rents are normally distributed.

(d) Do you think the rents are normally distributed? Keep in mind the normal distribution is symmetric about its mean (looks the same on both sides). If it isn't symmetric, what does it look like?

8. A national math competition advances to the second round only the top 5% of all participants based on scores from a first round exam. Their scores are normally distributed with a mean of 76.2 and a standard deviation of 17.1. What score, to the nearest whole number, would be necessary to make it to the second round? To start, look at the table and see if you can determine the z-value that corresponds to the top 5%.

STANDARD NORMAL DISTRIBUTION BASED ON Z-VALUE

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998



Name: ___________________________________ Date: ___________________

SAMPLING A POPULATION
COMMON CORE ALGEBRA II

The purpose of inferential statistics is to be able to learn something about a population by finding statistics about a sample from the population.

***Exercise* #1:** Why does it make sense to take a sample of a population instead of using the entire population to calculate statistics?

***Exercise* #2:** A population consisting of 200 adult males and females (100 of each) have heights in inches shown in the table and on the dot plot. The population mean is 67.0 inches and the population standard deviation is 6.3 inches (both rounded to the nearest tenth).

(a) This population has two peak heights, one at 65 inches and one at 70 inches. Statisticians would say that this data set is **bimodal**. Why do you think this population is **bimodal**?

(b) Why would a normal distribution **not** fit this population well?

***Exercise* #3:** Take a random sample of 40 of these heights. To do this, use a random integer generator on your calculator to generate numbers between 1 and 200 (the data point number in the shaded columns). Then, write down the value that goes with that data point number. You will certainly get repeated height values but if a data point number repeats, pick another one (circle the 40 you choose in the table). Write your sample values below.

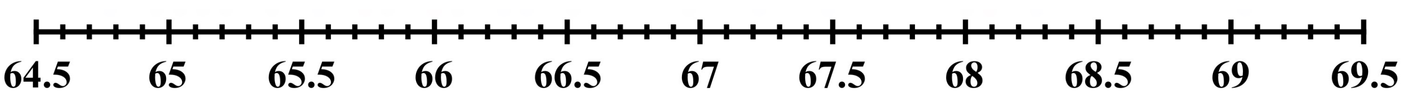
***Exercise* #4:** Calculate the sample mean, $\bar{x}$, and sample standard deviation, s, of your data set. Round both values to the nearest tenth.

***Exercise* #5:** We hope that a sample will fairly represent the overall population. If our sample is **representative** of our population then our sample mean will be close to our population mean and our sample standard deviation will be close to our population standard deviation.

(a) List your sample mean and the population mean below. Was your sample mean an overestimate or underestimate of the population mean?

(b) Do the same for the standard deviation.

***Exercise* #6:** Collect **at least** 20 sample means from fellow classmates (the more the better). Write the sample means below and plot their distribution on the dot plot shown.



Distribution of Sample Means

***Exercise* #7:** Find the mean of the sample means (the average of the averages) from Exercise #6. Find the sample standard deviation of the sample means from Exercise #6. Round both values to the nearest tenth.

***Exercise* #8:** How does the mean of the sample means compare with the population mean? Is it close or far away?

***Exercise* #9:** Is there more variation in the population or in the sample means from Exercise #6? Explain your answer.

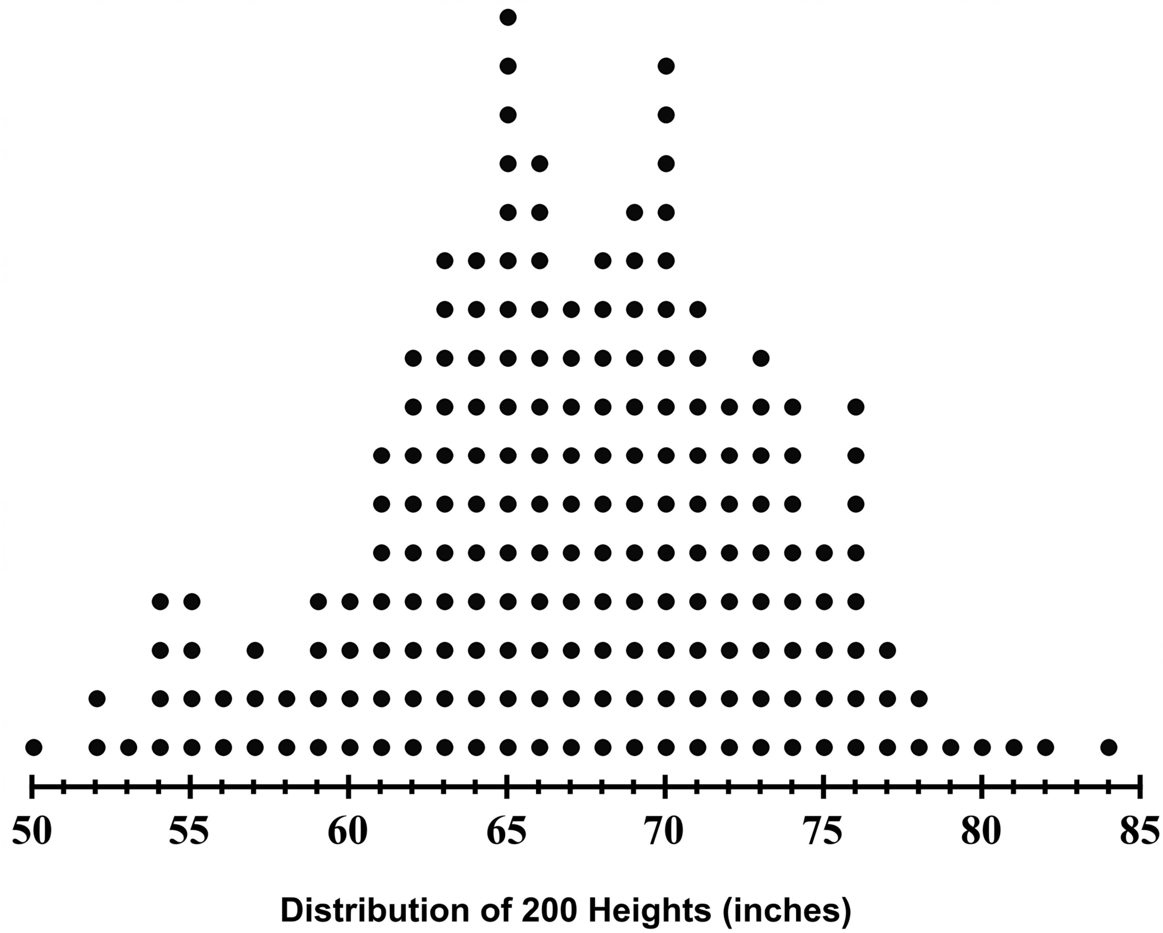
***Exercise* #10:** The **Central Limit Theorem** in statistics states that no matter how the population is distributed, if samples of size n are taken from the population with mean μ and standard deviation σ then the distribution of sample means will have the following characteristics:

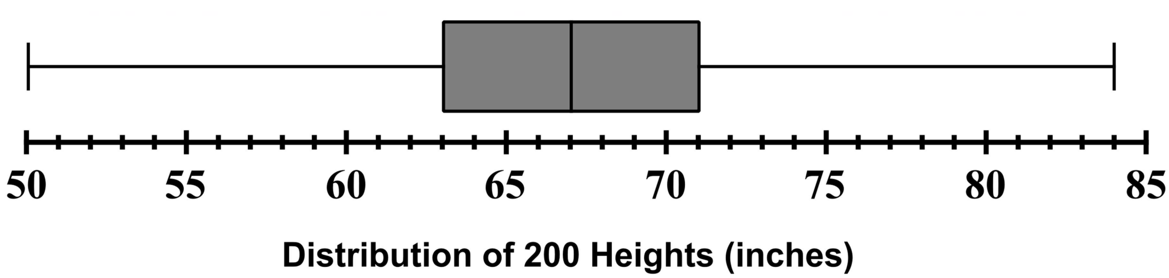
$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Does your sample of sample means from Exercise #6 support the Central Limit Theorem? Explain.

POPULATION OF 200 HEIGHTS IN INCHES
COMMON CORE ALGEBRA II

Data Point	Value	Data Point	Value	Data Point	Value	Data Point	Value	Data Point	Value
1	70	41	63	81	63	121	74	161	73
2	59	42	59	82	68	122	64	162	73
3	68	43	70	83	66	123	64	163	74
4	65	44	66	84	58	124	73	164	70
5	76	45	62	85	56	125	73	165	65
6	64	46	65	86	61	126	67	166	76
7	62	47	57	87	65	127	78	167	80
8	66	48	61	88	64	128	69	168	66
9	63	49	62	89	66	129	73	169	74
10	65	50	65	90	55	130	78	170	72
11	52	51	63	91	70	131	68	171	69
12	66	52	67	92	60	132	66	172	70
13	70	53	66	93	55	133	64	173	71
14	71	54	63	94	67	134	65	174	79
15	54	55	65	95	59	135	70	175	70
16	62	56	66	96	54	136	70	176	73
17	63	57	63	97	53	137	69	177	75
18	61	58	68	98	73	138	69	178	61
19	60	59	68	99	50	139	84	179	73
20	70	60	69	100	60	140	58	180	72
21	54	61	57	101	72	141	72	181	70
22	68	62	61	102	71	142	67	182	71
23	62	63	56	103	67	143	74	183	76
24	63	64	70	104	71	144	71	184	65
25	57	65	67	105	69	145	75	185	65
26	55	66	73	106	61	146	76	186	66
27	55	67	69	107	72	147	66	187	59
28	68	68	64	108	67	148	60	188	71
29	69	69	76	109	65	149	74	189	66
30	52	70	62	110	66	150	65	190	69
31	65	71	68	111	67	151	61	191	70
32	64	72	69	112	77	152	71	192	69
33	68	73	64	113	68	153	67	193	67
34	62	74	54	114	81	154	65	194	75
35	64	75	63	115	82	155	72	195	77
36	68	76	65	116	76	156	74	196	71
37	62	77	74	117	74	157	70	197	70
38	63	78	63	118	64	158	76	198	75
39	65	79	72	119	77	159	75	199	72
40	62	80	64	120	69	160	71	200	76





Population Parameters:

mean = $\mu = 67.0$ in

standard deviation = $\sigma = 6.3$ in

Name: ___________________________________ Date: _________________

Sample Means
Common Core Algebra II

The vast majority of the statistics that you've done so far has been **descriptive**. With descriptive statistics, we summarize how a data set "looks" with measures of central tendency, like the mean, and measures of dispersion, like the standard deviation. But, the more powerful branch of statistics is known as **inferential** where we try to **infer** properties about a population from samples that we take. We do this by using **probability** and **sampling variability** to estimate how likely the sample is given a certain population.

We begin our multi-lesson investigation into **inferential statistics** with the most basic question. How can we estimate the **population mean**, μ, if we know a sample mean, $\bar{x}$? Before answering this question, though, we need to investigate the distribution of sample means.

***Exercise* #1:** Say we are investigating the heights of 16 year old American males. Say we know that the population mean height is 65.3 inches with a standard deviation of 4.2 inches. Let's say we take a sample of 16 year old American males. The sample has a size of 30.

(a) Will the mean height of the sample always be 65.3? Why or why not? Could it be significantly different?

(b) Will the standard deviation of the sample be 4.2 inches? Would you expect more or less variation in a sample versus a population?

(c) Run the program NORMSAMP with a mean and standard deviation given above and a sample size of 30. Do at least 200 simulations (500 is preferable). State the minimum and maximum of the sample means, the mean of the sample means, and the standard deviation of the sample means.

min sample mean = _____________ Mean of means: Standard deviation of means:

max sample mean = _____________

(d) How does the **variability** of the sample means compare to the **variability** of the population? Is it more or less?

(e) State the mean of the sample standard deviations. How does it compare to the population standard deviation? Could you use the standard deviation of a sample to estimate the standard deviation of the population?

We can use our simulation to decide whether a sample could have come from a given population. We can even quantify how likely it would be to happen. This is known as establishing **confidence**.

***Exercise* #2:** Mr. Weiler took a sample of 30 16-year old males and found the mean height of the sample to be 66.4 inches. Do you believe this sample came from this population? Why or why not? Examine the results of your simulation. Quantify how likely this sample (or one greater) was to come from the population simulated.

Strangely enough, this process can be used in order to give a **confidence interval** for the population mean if we know the sample mean. This is important because in reality **the population mean is almost never known and is what we want to infer from the sample** mean. The next set of exercises will illustrate how this is done using simulation.

***Exercise* #3:** A sample of 50 ripe oranges were taken from a large orchard in order to estimate the mean weight of a ripe orange. The sample mean was 212 grams and the sample standard deviation was 34 grams.

(a) Why does it seem reasonable to use the sample standard deviation as an estimate of the population standard deviation? See Exercise #1(e).

(b) Run a simulation using the 212 as the population mean (even though it is the sample mean) and use 34 as the population standard deviation. State the 5^{th} percentile sample mean and the 95^{th} percentile sample mean.

5^{th} Percentile Sample Mean = _____________

95^{th} Percentile Sample Mean = ____________

(c) Now, try your simulation again, but use the 5^{th} percentile sample mean as the population mean. Where does the 212 lie on the distribution in terms of percentile? Notice how close this is to the 95^{th} percentile.

(d) Now, try your simulation again, but use the 95^{th} percentile sample mean as the population mean. Where does the 212 lie on the distribution in terms of percentile? Notice how close this is to the 5^{th} percentile.

(e) What both (c) and (d) tell us is that by using the 5^{th} and 95^{th} percentile values based on our original sample mean, we have actually found the **lowest possible population mean** and **highest possible population mean** that could have resulted in that sample mean 90% of the time. Write the 90% confidence interval below for μ based on (b).

Name: ____________________________________ Date: __________________

SAMPLE MEANS
COMMON CORE ALGEBRA II HOMEWORK

FLUENCY

1. The mean of the sample means is

(1) Greater than the population mean.
(2) Less than the population mean.
(3) Equal to the population mean.
(4) Could be greater or less than the population mean. _________

2. The variation within the sample means is

(1) Less than the variation within the population.
(2) More than the variation within the population.
(3) Equal to the variation within the population. _________
(4) Could be more or less than the variation within the population.

APPLICATIONS

3. A factory has machines that fill 12 ounce soda bottles repeatedly with an average volume of 12.2 ounces and standard deviation of 0.9 ounces. A new machine was installed and 30 bottles were sampled. It was found that they had an average volume of 11.8 ounces. We want to investigate whether this mean is significantly lower than the original population mean.

(a) Run NORMSAMP with a mean of 12.2 and a standard deviation of 0.9. Run 100 simulations. What percentile rank would you give the 11.8 ounces (this will vary based on the simulation)?

(b) Based on your findings from (a), can you conclude that this sample mean likely came from the same population or a different population with a lower mean? Explain.

(c) Use the sample mean of 11.8 ounces and a standard deviation of 0.9 ounces to generate the 90% confidence interval for the population mean of the new soda filling machine by simulation. Use a sample size of 30 and at least 100 samples to generate your interval. Round you lower and upper estimate for μ to the nearest hundredth.

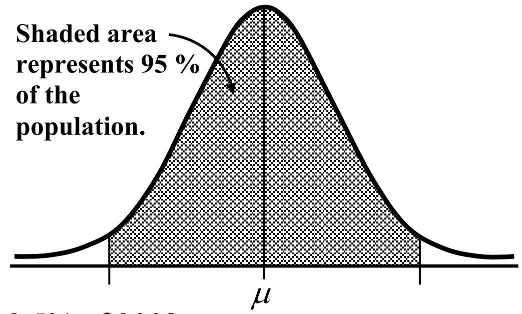
$\mu_L =$ ____________________

Interval: __________ $\leq \mu \leq$ __________

$\mu_H =$ ____________________

Let's work some now with the **95% confidence interval**. It will be easiest to use 200 simulations to generate this interval as we will soon see.

4. Consider a normal distribution with 95% of the probability (or distributions) **centered about the population mean** μ.



   (a) What percent lies in half of the shaded area shown in the diagram?

   (b) Explain why 2.5% of the population must lie in each of the un-shaded areas of the graph.

   (c) What is 2.5% of 200?

5. Researchers have found that the average number of hours per week spent by adults watching television is 34.2 with a standard deviation of 6.8 hours. Researchers wanted to determine if there was an effect to sampling people with tablets. They found that a random sample of 40 people with tablets had a sample mean of 36.3 hours per week with a sample standard deviation of 5.4 hours.

   (a) Run NORMSAMP with a population mean of 34.2 hours and a standard deviation of 6.8 hours. Do 200 simulations. This will take some time (approximately 8 minutes). Based on your results, what is the approximate percentile rank of 36.3 (remember it is out of 200 now)?

   (b) Do you have significant evidence that the 36.3 comes from a population with a mean that is higher than 34.2? Explain your thinking.

   (c) Now, let's attempt to construct the **95% confidence interval** for the sample whose mean was 36.3. Run a simulation with a mean of 36.3, a standard deviation of 5.4, a sample size of 40, and with 200 simulations. Find the 2.5th percentile as the lower limit and the 97.5th percentile as the upper limit.

   $\mu_L =$ ____________________

   Interval: __________ $\leq \mu \leq$ __________

   $\mu_H =$ ____________________

   (d) The **theoretical** (versus simulated) 95% confidence interval can be found using the formula below, where $\bar{x}$ is the observed sample mean and s is the sample standard deviation. Use this formula and compare to the interval from above.

   $$\bar{x} - 2 \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + 2 \cdot \frac{s}{\sqrt{n}}$$

   Interval: __________ $\leq \mu \leq$ __________

Name: ______________________________ Date: ______________

SAMPLE PROPORTIONS
COMMON CORE ALGEBRA II

Many times we are interested in determining a confidence interval for the population mean, μ, based on a sample mean, $\bar{x}$. Sometimes, though, we want to simply know what proportion, *p*, of a population shares a certain characteristic. We again infer characteristics about *p* based on the proportion of a sample, $\hat{p}$ (p "hat" as it is often called).

***Exercise* #1:** A school is trying to determine the proportion of students who own cell phones. They do a survey of **all** juniors and find that 168 out of 236 of have cell phones. They then take a **sample** of freshmen and find that 30 out of 52 freshmen in the sample own cell phones.

(a) Calculate the population proportion, *p*, of juniors who own cell phones. Round to the nearest hundredth.

(b) Calculate the sample proportion, $\hat{p}$, of freshmen who own cell phones. Round to the nearest hundredth.

Clearly, in the last example, the sample proportion of freshmen who own cell phones is less than the population proportion of juniors who own cell phones. But, can we attribute that variability to the two "treatments", i.e. juniors versus freshmen, or could the variability be due to **sampling variability**, i.e. the random chance that we just picked a group of freshmen who have an unusually low rate of cell phone ownership? We can establish how likely this is to happen by using simulation.

***Exercise* #2:** We would like to determine how likely it is that a sample of 52 out of a population with a proportion of cell phone ownership of 71% or 0.71 would result in a sample proportion of 0.58.

(a) Run the program PSIMUL with a *p* value of 0.71 and a sample size of 52 for 100 simulations. How many of the 100 simulations had a proportion less than or equal to 0.58?

(b) Based on your answer to (a), how likely is it that a sample of 52 from a population with a cell phone ownership of 71% would result in a sample proportion of only 0.58 or less?

(c) Is it possible that a sample of 52 from a population with a cell phone ownership rate of 71% could have a sample proportion of 0.58 or less? Justify your answer.

(d) What conclusion can you make about freshmen cell phone ownership compared to ownership by juniors? Explain.

Inferential statistics is never about proving beyond **any doubt** that a sample either can or cannot come from a certain population. It is about **quantifying how likely it is that it could come from a given population**. Let's continue exploring this question of sample proportions.

***Exercise* #3:** Let's say we have a population with a 0.25 proportion of being 65 years or older. Let's take different sized samples from this population and see how the sample proportions behave. Use the program PSIMUL to simulate a population with a proportion of 0.25 for various sample sizes and 100 simulations.

(a) Fill in the table below.

Sample Size	Low to High $\hat{p}$ values	Range in $\hat{p}$
10		
20		
50		
100		

(b) What was the effect of increasing the sample size on the sample proportions that were simulated? Why does this make sense?

(c) Run PSIMUL one more time with a sample size of 50 but for 200 simulations. Using your results, find the value that represents the 5^{th} percentile of $\hat{p}$ values. Find the result that represents that 95^{th} percentile of the $\hat{p}$ values. Then, write the **90% confidence interval** for this sample size coming from this population.

(d) If researchers surveyed 50 people walking out of a movie and found that 21 of them were 65 years or older, do you believe this sample came from the general population with a $p = 0.25$? Why or why not.

Name: ___________________________________ Date: _________________

SAMPLE PROPORTIONS
COMMON CORE ALGEBRA II HOMEWORK

APPLICATIONS

1. Historically, the proportion of emperor penguins with adult weights above 60 pounds is 0.64. Take this to be the population proportion for this characteristic.

   (a) A sample of 26 emperor penguins in a zoo found that 20 of the penguins had adult weights above 60 pounds. Calculate the sample proportion, $\hat{p}$, for this sample.

   (b) Run PSIMUL with a population proportion of $p = 0.64$ with a sample size of 26. Do 100 simulations. What percent of these simulations resulted in a $\hat{p}$ value at or above what you found in part (a)?

   (c) Do you have enough evidence from (b) to conclude that penguins raised in a zoo have a significantly higher proportion of weights above 60 pounds? Why or why not?

2. Let's stick with our emperor penguins from #1. Out of a sample of 56 penguins from a zoo, it was found that 43 penguins had weights over 60 pounds. Run PSIMUL again, but now with a sample size of 56. Continue to use $p = 0.64$ and 100 simulations. Do you now have stronger evidence that penguins raised in zoos have a higher proportion with weights over 60 pounds? Explain.

3. In general, as sample size increases, the range in the distribution of sample proportions

   (1) increases (3) decreases

   (2) stays the same (4) could increase or decrease _________

4. In a population with a proportion $p = 0.35$, if samples of size 30 were repeatedly taken, then we would expect approximately 90% of those samples proportions to fall within which of the following ranges?

   (1) 0.28 to 0.42 (3) 0.21 to 0.49

   (2) 0.18 to 0.54 (4) 0.31 to 0.39 _________

5. A sample of the graduating high school class was questioned about their plans after college. We worked with this sample of graduating seniors in our unit on probability. The two-way frequency chart below summarizes the results of the questionnaire. The school would like to investigate the effect of gender on the rate that students go to college.

(a) Calculate the sample proportion of students going to college for the subgroups male and female. Round to the nearest hundredth.

	Gender		
	Male	Female	Total
Going to College	16	13	29
Not Going to College	14	9	23
Total	30	22	52

$\hat{p}(\text{males}) = \dfrac{\text{number of males going to college}}{\text{total number of males}}$

$\hat{p}(\text{females}) = \dfrac{\text{number of females going to college}}{\text{total number of females}}$

The proportion for females going to college is higher than for males going to college (a greater percentage of females go to college than males). Is this due to **induced variability** or **sampling variability**? This boils down to asking if the difference is **statistically significant**.

(b) Design a simulation that would test how likely it is for a sample of 30 (the number of men) from a population that has the $\hat{p}(\text{female})$ would result in the $\hat{p}(\text{male})$ **or below**. Explain the simulation and what results you found.

(c) Based on the table above, would you conclude that the overall population proportion of females going to college is greater than the proportion of males? If you believe you have enough evidence from your simulation, explain why. If you do not believe you do, also explain.

(d) Why could this study be an example of both a sample survey and an observational study? Look back at their definitions from the first lesson to fully explain your answer. Also explain how the types of variability introduced.

Name: ____________________________________ Date: __________________

THE DIFFERENCE IN SAMPLE MEANS
COMMON CORE ALGEBRA II

In a classic **experiment** two or more **treatment groups** are randomly created and then subjected to the **different treatments**. The question then is how to determine if the variability seen between the two groups is due to the treatment.

***Exercise* #1:** Suppose 50 people were chosen to try out a new diet pill to help increase weight loss. The people are randomly divided into two groups. One is given the pill while the other is given a **placebo** (a pill designed to look the like real one, but with no medicine). There are two main ways variability can be introduced to the results. Discuss each.

Induced (Variability created because of the treatment the subject was placed in):

Natural (Variability just because people, animals, plants, etcetera, are naturally different):

The question, then, is how we can distinguish between the two types. Let's look at a case study.

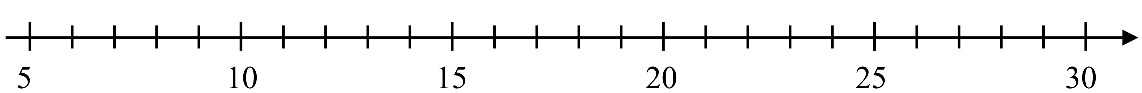
***Exercise* #2:** A seed company is trying to determine the effect of synthetic nutrients versus organic nutrients on the growth rate of corn plants. They select 40 seeds and randomly distribute the seeds to two groups of 20. The seeds in Group 1 are given the organic nutrients and the seeds in Group 2 are given the synthetic nutrients. After three weeks each plant's growth is measured in centimeters.

Group 1 (Organic): 6, 8, 10, 12, 12, 12, 13, 13, 14, 14, 16, 16, 17, 17, 18, 18, 20, 20, 22, 25

Group 2 (Synthetic): 9, 11, 12, 12, 15, 15, 15, 16, 17, 18, 18, 18, 19, 19, 19, 21, 21, 22, 24, 28

Enter these two lists in your calculator. State the mean of each. Then, create a box plot for each using the grid below. You may want to summarize the information you need under each heading.

Group 1 Group 2



***Exercise* #3:** Let's look at the **descriptive statistics** we have so far: the sample means and the box-plot. What does this data suggest? Does it give a clear indication that one treatment resulted in greater plant growth?

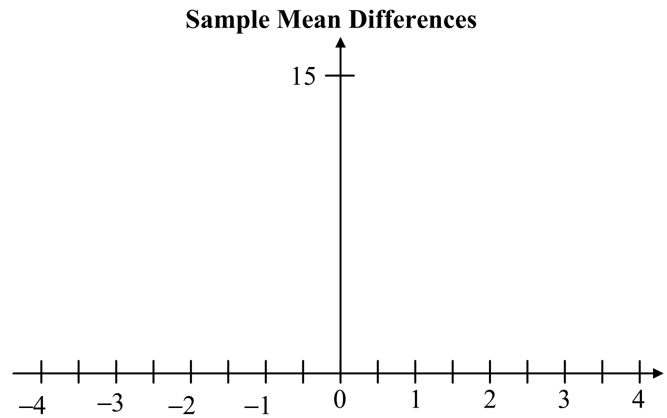
There are very sophisticated techniques to probabilistically determine what portion of the variability in these two data sets is due to **natural causes** and what is **induced**. But, we can run a simple simulation which can give us a very good sense. Consider just the question of the **difference in the sample means**. The program MEANCOMP will take our two groups of data and randomly scramble them up into two new groups. It will do that over and over again and calculate the difference in the means of the groups.

***Exercise* #4:** If $\overline{x_1}$ represents the mean of Group 1 and $\overline{x_2}$ represents the mean of Group 2, do the following.

(a) Find the **observed difference** in the sample means:

$\overline{x_2} - \overline{x_1} =$

(b) Run the program MEANCOMP with 100 simulations. Use your calculator to create a frequency histogram on the axes below for the sample mean differences. Point out where on the histogram the **observed difference** falls.



(c) Look at the data list containing the sample mean differences. Given there are 100 differences, what percent of the differences were at or above the observed difference?

(d) How confident are you about the **observed difference in sample means** being due to **induced variability** and not **natural variability**? Justify.

Name: ___________________________________ Date: _________________

THE DIFFERENCE IN SAMPLE MEANS
COMMON CORE ALGEBRA II HOMEWORK

APPLICATIONS

1. In an experiment, two main types of variability are introduced. Explain how both affect the results of the experiment. Give examples to support your descriptions.

   Induced Variability   Natural Variability

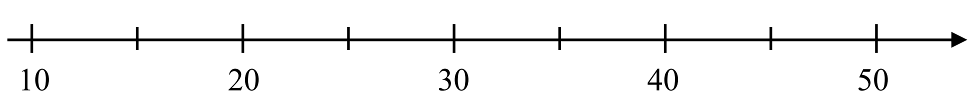
2. A simulator takes the data from the various treatments, randomly scrambles them together to create groups that contain mixed treatments. Explain how this helps quantify the question of natural variability.

3. Researchers in a sleep lab at a college decide to see how a night of no sleep affected the ability of volunteers to answer 50 addition problems in a minute time span. Thirty volunteers were randomly assigned to two groups. Group 1 was not allowed sleep and Group 2 slept normally. Their results, in terms of questions answered out of 50, are given below. As in the lesson, find the sample means and graph a box plot for each.

   Group 1: 11, 14, 16, 17, 23, 25, 25, 27, 30, 31, 33, 34, 34, 36, 38

   Group 2: 18, 22, 24, 25, 30, 30, 32, 33, 34, 34, 36, 37, 42, 44, 48

   Group 1   Group 2



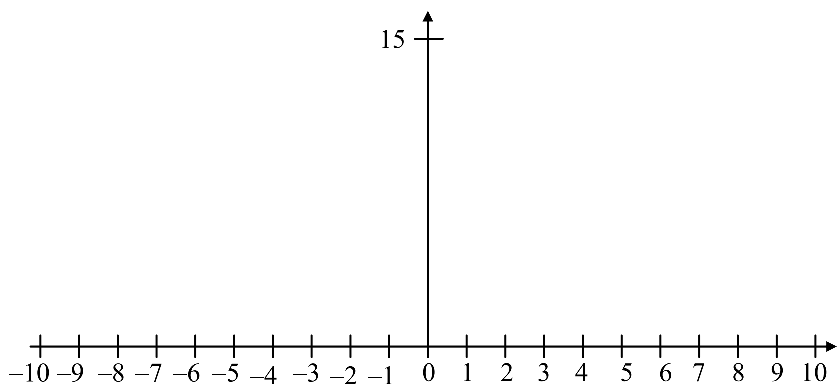
4. Does it appear that getting sleep helps in the ability to answer addition problems? What descriptive statistics can you use to strengthen your argument?

5. Is it true that a person who gets sleep will always answer more addition problems than a person who has not gotten any sleep? Support your answer from the experimental results.

6. Run the program MEANCOMP with 100 simulations on these two data sets. Using your calculator, create a frequency histogram for the sample mean differences on the axes below. Mark on the distribution where the **observed difference** in the sample means lies.

Observed Difference:

$\overline{x}_2 - \overline{x}_1 =$



7. What percent of the simulated differences were greater than or equal to our observed differences? Show your calculation below.

8. Can we **confidently** conclude that the variability in sample means is due to the **treatment** or due to **natural variability**? Support your argument using the distribution above and your answer to 7.

Name: ____________________________________ Date: __________________

THE DISTRIBUTION OF SAMPLE MEANS
COMMON CORE ALGEBRA II

In order to be able to make **inferences** about population parameters based on sample statistics we first must understand how sample statistics, like the sample mean and sample proportion, are distributed. For example, if we take many samples from a population, how will the means of those samples distribute? In this lesson we will investigate this both with simulation and formally.

***Exercise* #1:** Using the Normal Distribution simulator, run a simulation for a population with a mean of $\mu = 50$ and a standard deviation of $\sigma = 15$ for a sample size of 30. Run 100 simulations.

(a) Does the distribution of sample means appear normal (i.e. like a normal distribution)? Explain.

(b) What is the mean of the sample means, symbolized by $\mu_{\bar{x}}$?

(c) What is the standard deviation of the sample means, symbolized by $\sigma_{\bar{x}}$, rounded to the nearest tenth? How does it compare with the standard deviation of the population?

(d) Based on this simulation alone, how likely would it be that a sample of this size taken from this population would have a mean greater than 2 standard deviations, $\sigma_{\bar{x}}$, above the mean?

THE CENTRAL LIMIT THEOREM

When a sample size is fairly large, say 30 or more, then the **distribution of all sample means** of a given size *n* will be **normally distributed** with:

1. **A mean**: $\mu_{\bar{x}} = \mu$ (the mean of the sample means will just be the mean of the population)
2. **A standard deviation**: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (the variation of sample means is smaller than that of the population)

***Exercise* #2:** Do the results of your simulation agree with the Central Limit Theorem. Explain.

***Exercise* #3:** The mean height of adult American males is 177 cm with a standard deviation of 7.3 cm. What is the standard deviation of the distribution of samples means from this population with a sample size of 50?

(1) 0.15 (3) 3.54

(2) 1.03 (4) 4.72

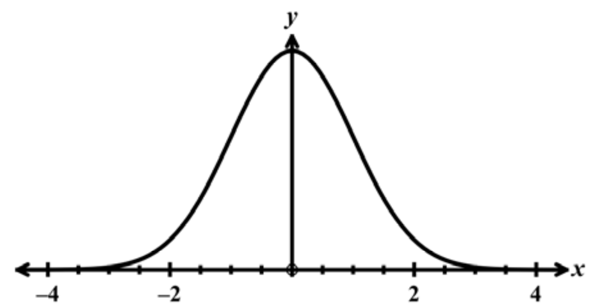
____

Because the distribution of sample means follows a normal distribution, we can determine how likely a sample mean would be given population parameters.

***Exercise* #4:** Jumbo eggs have a mean weight of 71 grams and a standard deviation of 3 grams. A sample of three dozen jumbo eggs was taken at a local egg processing plant and found to have a mean weight of 70 grams. Should this be of concern? Let's explore this question.

(a) What is the standard deviation of sample means of this size from the described population?

(b) What is the z-score for this particular sample mean? Illustrate this on the standard normal curve shown below.



(c) Using either tables or your calculator, determine the probability that a sample of this size would have a mean of 70 grams **or lower**. Round to the nearest tenth of a percent. Shade this area on the normal graph shown in (c).

(d) Why does it make sense in part (c) to determine the probability of having a sample with 70 grams or lower? What does the probability from part (c) tell you?

***Exercise* #5:** Given a population with a mean of 58 and a standard deviation of 12, which of the following represents the probability of getting a sample mean of 61 or greater with a sample size of 50? Show the analysis that leads to your choice.

(1) 7.2% (3) 18.0%

(2) 3.9% (4) 24.2%

______

***Exercise* #6:** In a normal distribution, approximately 95% of all data lie within two standard deviations of the mean. This includes normal distributions of sample means. If a population has a mean of 130, a standard deviation of 8 and samples of size 30 are taken, find the sample mean two standard deviations below the mean and two standard deviations above. Round both means to the nearest hundredth.

Name: ____________________________________ Date: __________________

THE DISTRIBUTION OF SAMPLE MEANS
COMMON CORE ALGEBRA II HOMEWORK

FLUENCY

1. For each of the population standard deviations given, calculate the standard deviation of the sample means, i.e. $\sigma_{\bar{x}}$, with a sample size of the specified value of n. Show your calculation. Round all answers to the nearest tenth.

(a) $\sigma = 12.5$ and $n = 40$

$\sigma_{\bar{x}} =$

(b) $\sigma = 22$ and $n = 65$

$\sigma_{\bar{x}} =$

(c) $\sigma = 2.7$ and $n = 127$

$\sigma_{\bar{x}} =$

(d) $\sigma = 35$ and $n = 237$

$\sigma_{\bar{x}} =$

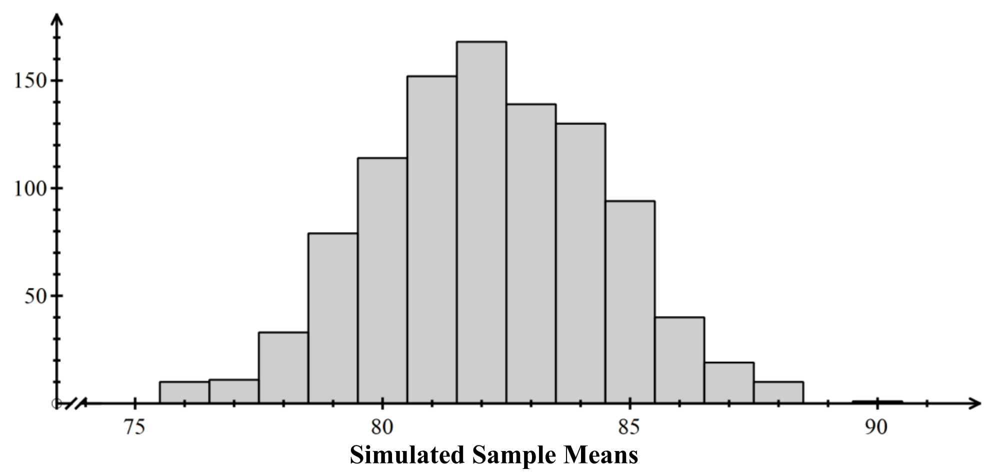
(e) $\sigma = 23.8$ and $n = 30$

$\sigma_{\bar{x}} =$

(f) $\sigma = 18.5$ and $n = 35$

$\sigma_{\bar{x}} =$

2. A simulation is done using a population with a mean of 82, a standard deviation of 15 and samples of size 40. When 1000 samples are simulated, the distribution of simulated sample means is created and shown. The mean of these sample means is 82.09 and the standard deviation is 2.33. Does this simulation support the conclusions of the Central Limit Theorem? Explain.



3. On a standardized test, the standard deviation of the scores was 14.8 points. If samples of size 60 were taken from this test's results, which of the following would be closest to the standard deviation of the means of these samples?

(1) 0.25
(3) 1.91
(2) 0.48
(4) 2.29

____

APPLICATIONS

4. In 2014, new cars had an average fuel efficiency of 27.9 miles per gallons with a standard deviation of 6.8 miles per gallon. A sample of 30 new cars is taken.

   (a) What is the probability the sample has a mean gas mileage between 27 and 29 miles per gallon?

   (b) What is the probability that the sample has a mean gas mileage greater than 30 miles per gallon?

   (c) If a sample of 30 trucks had a sample mean gas mileage of 24.3 miles per gallon, why is it reasonable to assume that all trucks have a lower overall gas mileage than cars? Explain

5. The average length of a cell phone call in 2012 was 1.80 minutes with a standard deviation of 0.32 minutes. A sample of 50 cell phone calls made by users less than 20 years old was taken and had a mean call length of 1.89 minutes.

   (a) What is the probability that a sample of 50 from a population with a mean of 1.80 and a standard deviation would have a mean call length of 1.89 minutes or longer.

   (b) What conclusion can you make about the average phone call length of users younger than 20 compared to the general population? Explain.

REASONING

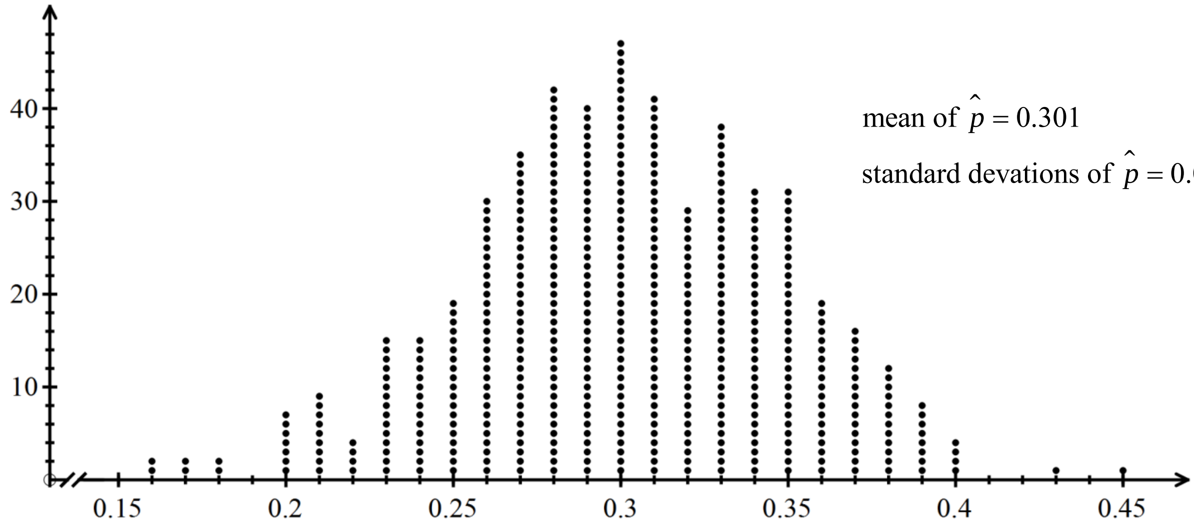
6. A population has a standard deviation of 37. If a researcher is designing a study so that the distribution of sample means has a standard deviation of less than 5, what is the smallest sample size that can be used?

Name: ______________________________________ Date: ___________________

THE DISTRIBUTION OF SAMPLE PROPORTIONS
COMMON CORE ALGEBRA II

In the last lesson we saw how the **distribution of sample means** was **normal**. The **Central Limit Theorem** allowed us to find the standard deviation of these sample means. In this lesson, we will look at the same phenomena with sample proportions.

***Exercise* #1:** A simulation of samples taken from a population with a proportion, *p*, of 0.3 was created. The simulation had a sample size of 100 and 500 simulations were run. The sample proportions, $\hat{p}$, were calculated and their distribution is shown below:



(a) What does the shape of this distribution resemble? Explain.

(b) What is true about the mean of the sample proportions?

The distribution of sample proportions is governed by a very similar phenomena to the distribution of sample means via **The Central Limit Theorem**. The characteristics of the distribution are given below.

THE DISTRIBUTION OF SAMPLE PROPORTIONS

The distribution of sample proportions, $\hat{p}$, from a population with a proportion *p* and a sample size of *n* will:

1. Approximate a normal distribution

2. Have a mean of the population proportion, *p*.

3. Have a standard deviation given by $\sqrt{\frac{p(1-p)}{n}}$

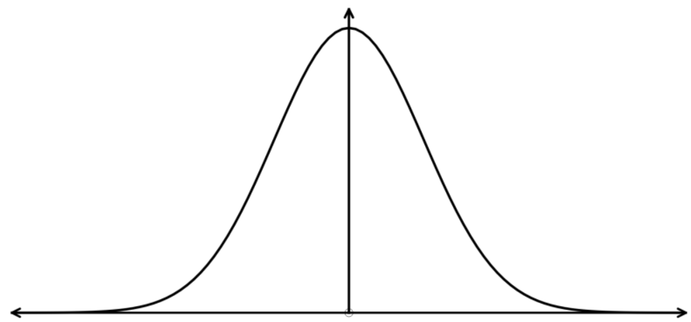
***Exercise* #2:** Does the standard deviation from the simulation agree with that predicted with the above formula?

Since sample proportions will be **normally distributed**, we can perform calculations similar to those done for **sample means**. In other words, we can see how likely a range of sample proportions would be given a particular population proportion.

***Exercise* #3:** Suppose the percent of seniors in high school that own a cell phone is 82%. If a random sample of 50 high school seniors was taken, determine the following:

(a) The standard deviation of sample proportions for this population proportion given this sample size. Show the calculation that leads to your answer.

(b) The probability that the sample proportion will be within 3% of the 82% proportion. Illustrate your work on the general normal curve below.



(c) Find each of the following probabilities. Round each answer to the nearest tenth of a percent.

(i) the sample proportion will be less than 75%

(ii) the sample proportion will be greater than 95%

***Exercise* #4:** Political polls can be tricky. Let's say that 47% of the public will vote for a particular candidate in the upcoming election. If a newspaper takes a random poll of 200 voters, what is the probability that this sample will have a proportion larger than 50%, thus predicting a win for this candidate?

Name: ____________________________________ Date: __________________

THE DISTRIBUTION OF SAMPLE PROPORTIONS
COMMON CORE ALGEBRA II HOMEWORK

FLUENCY

1. For each of the following population proportions, p, find the standard deviation of the sample proportions, $\hat{p}$, given the sample size n. Show your calculation. Round to three decimal place accuracy (nearest thousandth).

(a) $p = 0.34$ and $n = 50$

$\sigma_{\hat{p}} =$

(b) $p = 0.5$ and $n = 400$

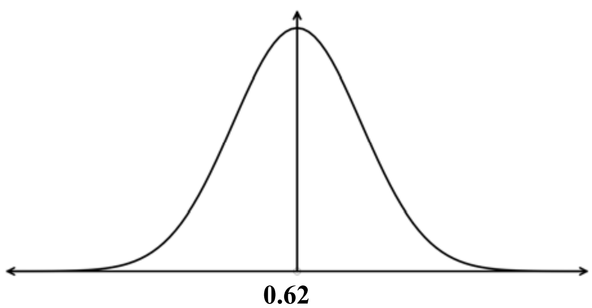
$\sigma_{\hat{p}} =$

(c) $p = 0.25$ and $n = 100$

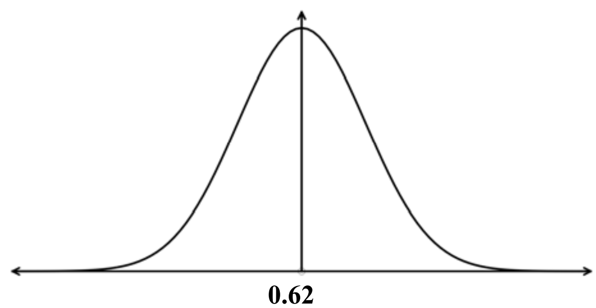
$\sigma_{\hat{p}} =$

2. A population has a proportion of 0.62. A sample of size 40 was taken from this population. Determine the following probabilities. Illustrate each on the normal curve shown below each part.

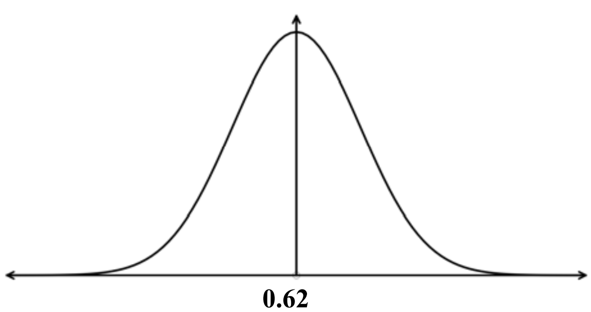
(a) The probability the sample has a proportion between 0.5 and 0.7.



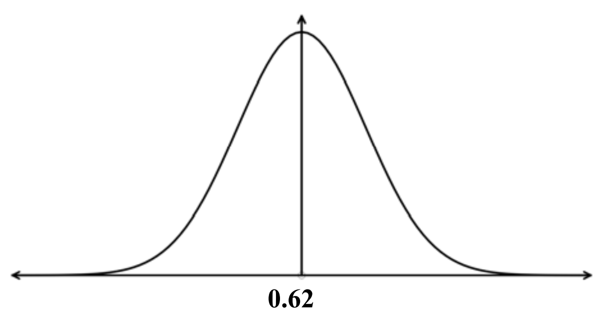
(b) The probability the sample has a proportion within 5% of the population proportion.



(c) The probability that the sample has a proportion less than 0.50.



(d) The probability that the sample has a proportion greater than 0.80



APPLICATIONS

3. A candidate for political office has support from 40% of the public. If a random sample of 100 members of the public was taken, which of the following is closest to the probability that the sample had a proportion of 50% or greater support for this candidate?

(1) 2%   (3) 14%

(2) 7%   (4) 24%

____

4. A school will offer pizza on Friday's if at least 30% of the students will buy it. A sample of 50 students are asked if they would buy pizza on Friday and 10 respond that they would.

(a) Determine the probability of getting a sample of this size with the proportion or lower given a population with a proportion of 0.30.

(b) Should the school offer pizza on Fridays? Explain your choice by reflecting on what your answer from part (a) tells you.

5. If a 45% of a population likes a particular soda, then what range below shows all sample proportions within two standard deviations of the population proportion if the samples have a size of 70?

(1) 38% to 52%   (3) 33% to 57%

(2) 20% to 70%   (4) 28% to 62%

____

REASONING

6. Juniors at a high school own internet enabled devices at a rate of 71%. If 52 freshmen were sampled and only 58% of them owned internet enabled devices, is this enough proof to state that freshmen own these devices at a lower rate than juniors? Explain based on probability.

Name: ___________________________________ Date: _________________

MARGIN OF ERROR
COMMON CORE ALGEBRA II

In **inferential statistics** we attempt to **infer** characteristics about a **population** from **sample statistics**. But, these inferences always have a degree of uncertainty in them. Many times, this uncertainty itself is quantified by a **margin of error**, which is a measurement of how accurate we believe our sample statistic to be relative to the population.

***Exercise* #1:** A recent poll found that 36% of all respondents would vote for Candidate A in an election. The poll reported a **margin of error** of 4%. Give an interpretation of what this margin of error means in terms of the 36% support for Candidate A.

The **margin of error** allows us to give a range of values that are reasonable for a population parameter based on a sample statistic. We will only consider what is known as the **95% margin of error**. Although, there are many others based on **confidence level**.

THE 95% MARGIN OF ERROR

The **uncertainty level** that would guarantee a 95% chance that the population parameter falls within a certain range of values.

***Exercise* #2:** In any normal distribution, how much of the data falls within two standard deviations of the mean? Do this quickly using your calculator.

Since roughly 95% of all normally distributed data fall within two standard deviations of the mean, we use two **standard deviations** to develop the **margin of error**. The next exercise will illustrate how this is done for a population proportion.

***Exercise* #3:** In a sample survey, 50 people were randomly sampled about their favorite soda. If 38% of them listed Soda A as their favorite, then answer the following questions.

(a) Based on a proportion of $p = 0.38$, what is the standard deviation of sample proportions of this size, $\sigma_{\hat{p}}$?

(b) What is the margin of error for this survey? What would be an acceptable range of values for the population proportion?

Margins of error are commonly found in surveys and other types of studies that are trying to determine a population proportion based on a sample proportion.

***Exercise* #4:** In a poll of 500 potential voters, Candidate A led Candidate B by a 46% to 39% margin. Could these two candidates actually be tied in the population as a whole? Justify your response.

The **margin of error** can also be useful in working with **sample means**, which will also have a **normal distribution**.

***Exercise* #5:** If a sample of three dozen jumbo eggs had a mean weight of 69.7 grams and a sample standard deviation of 3.2 grams, answer the following question.

(a) What would be a reasonable estimate for the standard deviation of the sample means, i.e. $\sigma_{\bar{x}}$?

(b) Based on (a), what would the margin of error for the population mean weight be?

(c) Jumbo eggs are considered to be eggs with weights at or above 70 grams. Is this within the margin of error for this sample? Explain.

***Exercise* #6:** In 2015, a survey of fifty 20 to 24 year olds was done to determine their mean weekly earnings. The survey found a sample mean of $495 with a standard deviation of $48. If the *World Almanac* reported the 2014 mean weekly earnings of this age range to be $472, do the results of this survey conclusively imply an increase in the mean weekly earning from 2014 to 2015? Explain.

Name: ____________________________________ Date: __________________

Margin of Error
Common Core Algebra II Homework

Fluency

1. Use the two standard deviation rule to determine the margin of error, to the nearest thousandth, for each of the following proportions with the given sample size. Show the work that leads to your answer.

(a) $p = 0.35$ and $n = 40$

(b) $p = 0.72$ and $n = 100$

(c) $p = 0.5$ and $n = 50$

(d) $p = 0.25$ and $n = 30$

2. Assuming a population characteristic has a standard deviation, σ, of 38. Calculate the margin of error on the population mean given a sample of each of the sizes given below. Show how you calculate your answer.

(a) $n = 30$

(b) $n = 100$

(c) $n = 1000$

(d) Generally, as the sample size increases, what happens to the margin of error? Why do you think this occurs?

(e) What is the minimum sample size needed for the margin of error to be 2 or less? Show or explain how you determined your solution.

APPLICATIONS

3. In an election poll, 200 people were surveyed and 45% expressed their likelihood to vote for a particular candidate. The margin of error on this estimated support is closest to

(1) 2% (3) 7%

(2) 3% (4) 12%

____

4. In a survey of 125 students, 64% of them preferred to start the school day an hour later.

(a) Calculate the margin of error for this survey to the nearest tenth of a percent. Show your work.

(b) The administration of the school will only continue to study the feasibility of starting the day later if there is at least 70% support amongst students. Does this fall within the margin of error of the survey?

5. A consumer group is trying to determine the mean amount that a family of four spends on food per week. They perform a phone survey of 300 random families of four and find a sample mean of $241.50 with a standard deviation of $46.72.

(a) Determine an estimate for the standard deviation of the sample means, $\sigma_{\bar{x}}$. Show your calculation.

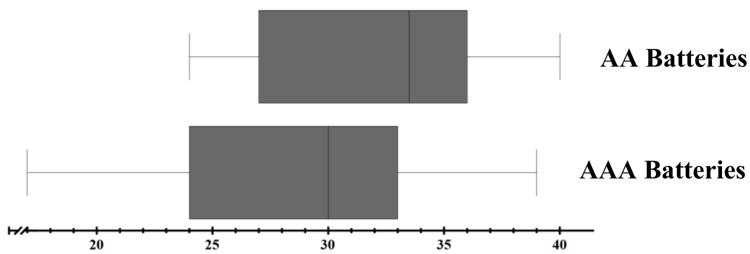
(b) What is the margin of error for the mean amount spent on food per week?

(c) If the *World Almanac* found that the mean amount spent by all four person families in 2015 was $244.90, was this within the margin of error you found in (b)? Explain or show how you arrived at your conclusion.

Name: ____________________________________ Date: _______________

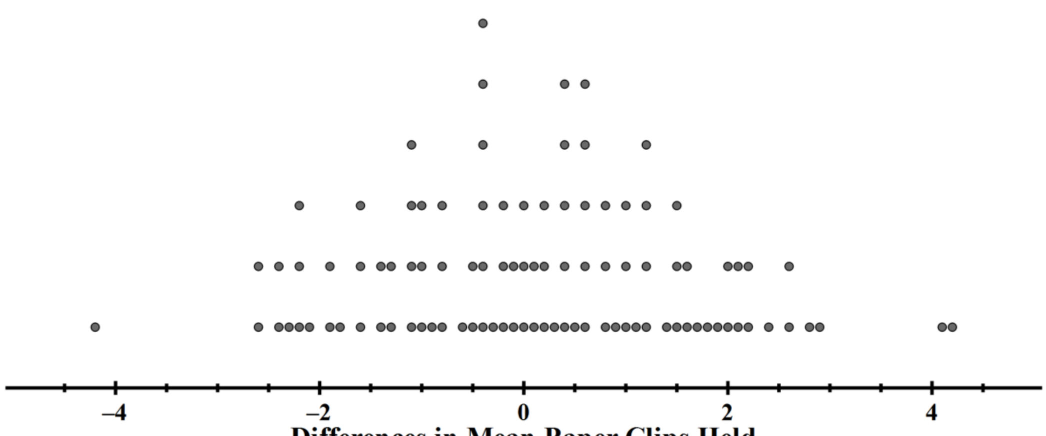
STATISTICAL SIMULATION PACKET
COMMON CORE ALGEBRA II

1. For his science fair project, Max hypothesized that AA batteries create a more powerful electromagnet than AAA batteries. He tested this by hooking up 30 of each type of battery and testing how many paperclips the electromagnet could hold. He found that the 30 AA batteries held on average 3.8 more paper clips than the 30 AAA batteries. The box plot distribution of the two trials is shown below.



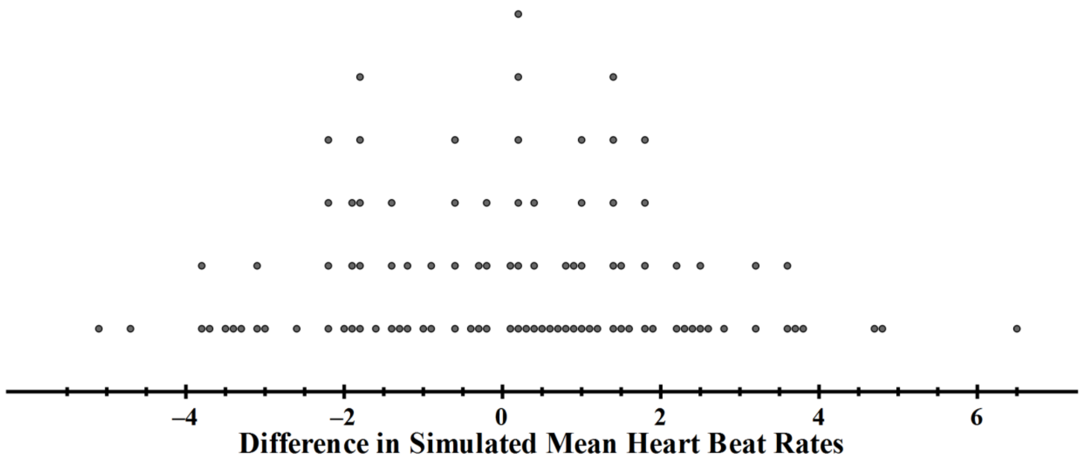
Based on the box plots, explain why it would be incorrect for Max to conclude that AA batteries will always hold more paper clips than AAA batteries.

To determine if the observed difference in the trial means was significant, he randomized the results of the 60 batteries 100 times and calculated the differences in the mean paper clips held for each simulation. The results are shown below.



Explain why, based on these simulations, that Max can conclude that there is statistically significant evidence to support his hypothesis that AA batteries hold more paperclips than AAA batteries.

2. A study was done to determine if regular consumption of coffee had an effect on the resting heart rate of humans. In the study, the resting heart rate of 25 adults who consume coffee and the resting heart rate of 25 adults who don't drink coffee was measured and compared. It was found that the participants in the study who consume coffee had an average heart beat of 3.2 beats per minute faster than those who did not drink coffee. To determine if this difference in means was significant, the 50 data values were randomly assigned to two groups for a total of 100 simulations and the differences in the sample means were calculated. Those results are shown below.



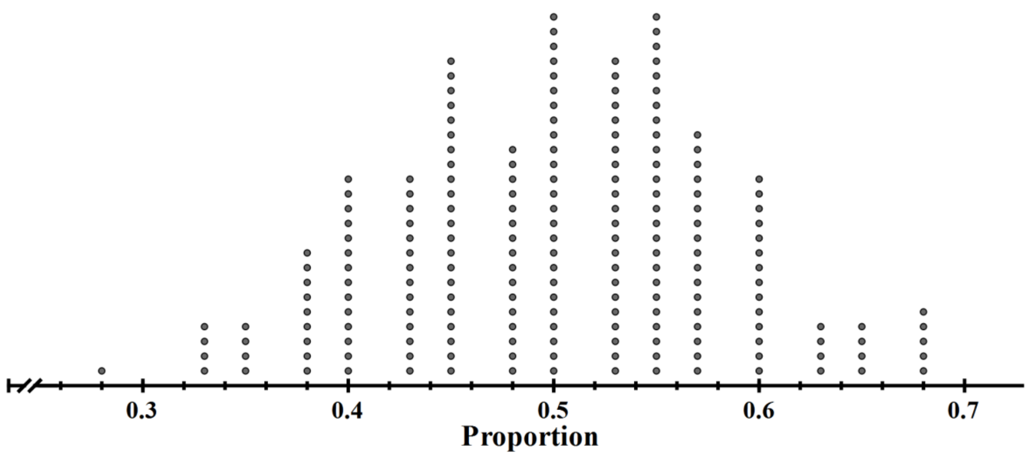
Give an argument for why the observed difference in mean heart rates is not significant based on the simulation results.

Based on the simulation results, give a difference in trial means that would have been significant. Explain your choice.

3. In order to spread news faster, a school would like to start using a common social network app. They will only do so if at least 50% of the student body regularly uses the app. They survey 40 students and find that 18 out of the 40 regularly use the app.

What is the sample proportion for this survey? Show your calculation.

The school still believes there is a reasonable chance that at least 50% of the student body regularly uses the app. They run a simulation of 200 more surveys of 40 students assuming that 50% of the students use the app. The simulation results are shown below.



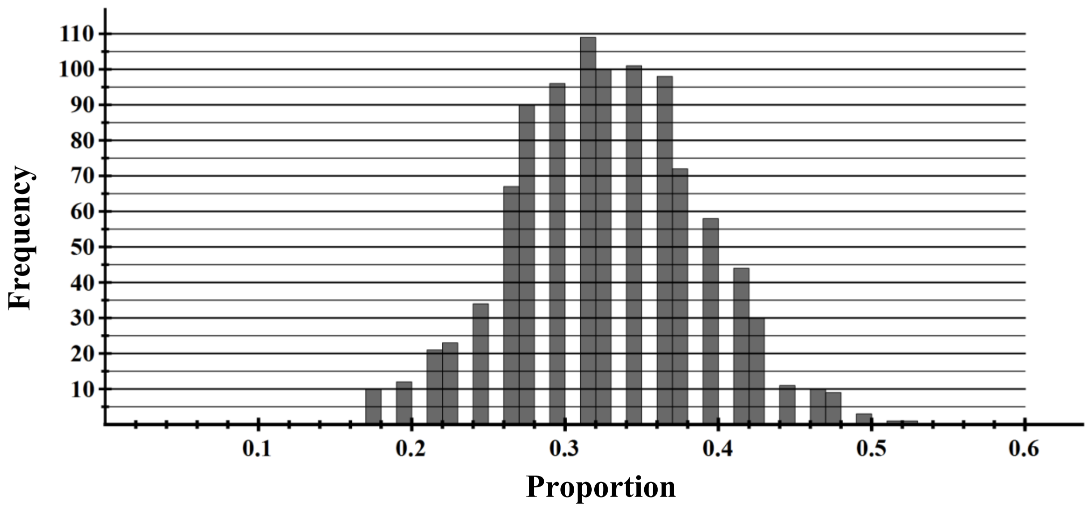
Assuming a 95% confidence level, give an estimate for the margin of error for this simulation. Explain your choice.

Is there evidence from this simulation and the survey that the population uses the app at a rate of at least 50% rate? Explain.

4. The Center for Disease Control and Prevention (CDC) recommends that adults get 7 or more hours of sleep per night. They estimate that 1 in 3 adults do not meet this requirement. A study was done to see if mothers of teenage children were more sleep deprived than the population as a whole. A survey of 60 mothers of teenage children found that 24 out of 60 got less than 7 hours of sleep per night.

What is the sample proportion of sleep deprived mothers of teenage children?

To determine if this sample proportion was significantly different than the population as a whole, researchers ran 1000 simulations with a population proportion of 0.33 and a sample size of 60. The results are shown below.



Explain why the simulation results do not provide statistically significant evidence that mothers of teenage children are more sleep deprived than the adult population on the whole? Explain.

Another survey of 60 fathers of newborns found that 31 of them got less than 7 hours of sleep per night. Do the simulation results now show that fathers of newborns are more sleep deprived than the population as a whole? Explain.

Name: ___________________________________ Date: _________________

LINEAR REGRESSION AND LINES OF BEST FIT
COMMON CORE ALGEBRA II

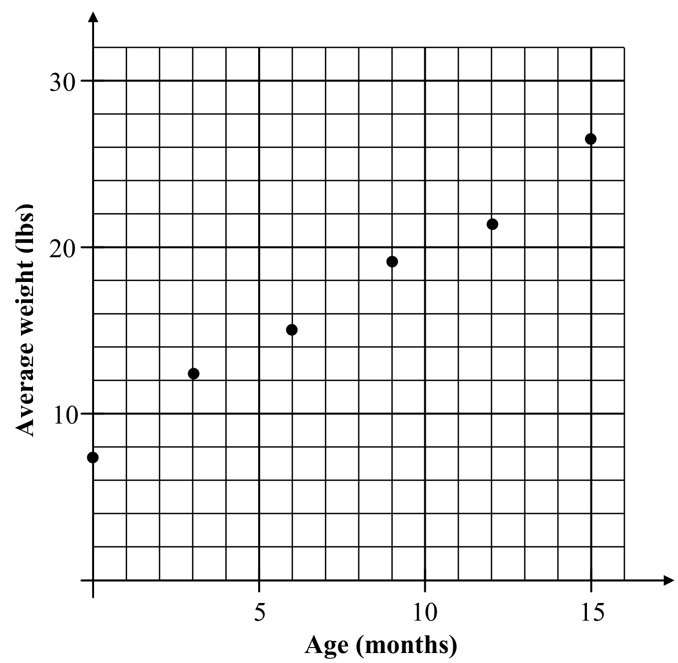
Oftentimes in science, a mathematical relationship between two variables is desired for predictive purposes. In the real world, the relationship between two variables is not always a perfect one, thus we often look for the "best" curve that can fit the data. Today we will review how to do this with a linear function.

***Exercise* #1:** A pediatrician would like to determine the relationship between infant female weights versus age. The pediatrician studies 100 newborn girls and finds their average weight at the end of 3 month intervals. The data is shown below and graphed on the scatter plot.

Age (months)	0	3	6	9	12	15
Average Weight (pounds)	7.2	12.2	15.1	19.4	21.5	26.3

(a) Using a ruler, draw a line that you think best fits this data. As a general guideline, try to draw it such that there are as many data points above the line as below it.

(b) By picking two points that are on the line (not necessarily data points), determine the equation of your best fit line. Round your coefficients to the nearest *tenth*.



(c) Using the linear regression command on your calculator, find the equation of the best fit line

(d) Use your calculator to determine the **linear correlation coefficient.** Round to the nearest *thousandth*. How can you interpret this value in terms of the variation in weight due to age?

***Exercise* #2:** Using the equation that your calculator produced in Exercise #1, predict the weight of a baby girl after 10 months. Round your answer to the nearest tenth of a pound.

The use of a model to predict outputs when the input is within the range of the known data is called **interpolation**. Interpolation tends to be fairly accurate.

***Exercise* #3:** Using the equation that your calculator produced in Exercise #1, predict the weight of a baby girl after 2 years. Round your answer to the nearest tenth of a pound.

The use of a model to predict outputs when the input is outside of the range of the known input data is called **extrapolation**. Models are most helpful when they can be used to extrapolate, but tend to be less accurate.

***Exercise* #4:** Biologists are trying to create a least-squares regression equation (another name for best fit line) relating the length of steelhead salmon to their weight. Seven salmon were measured and weighed with the data given below.

Length (inches)	22	24	28	34	39	42	48
Weight (pounds)	3.43	4.46	7.08	14.21	22.19	31.22	35.67

(a) Determine the least-squares regression equation, in the form $y = ax + b$, for this data. Round all coefficients to the nearest *hundredth*.

(b) Using your equation from part (a), determine the expected weight of a salmon that is 30 inches long.

(c) Using your equation from part (a), determine the expected weight of a salmon that is 52 inches long.

(d) In which part, (b) or (c), did you use interpolation and in which part did you use extrapolation? Explain.

Name: ___________________________________ Date: _________________

LINEAR REGRESSION AND LINES OF BEST FIT
COMMON CORE ALGEBRA II HOMEWORK

FLUENCY

1. Which of the following linear equations would best fit the data set shown below?

(1) $y = 2.4x + 18.7$ (3) $y = -1.6x + 27.2$

(2) $y = -0.8x + 18.1$ (4) $y = 1.9x - 15.6$

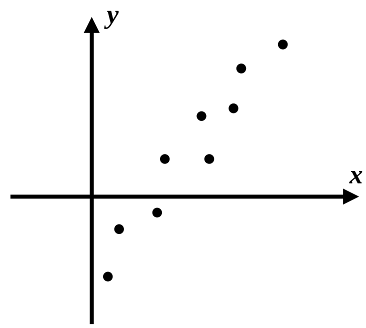
x	2	5	9	15
y	26	17	12	4

____

2. A scatter plot is shown below. Which of the following *could* be the equation of the best fit line for the data set?

(1) $y = 1.8x - 3.2$ (3) $y = -2.9x + 8.3$

(2) $y = -3.5x - 12.4$ (4) $y = 6.5x + 3.9$



____

3. A line of best fit was created for a data set that only included values of x on the interval $12 \le x \le 52$. For which of the following values of x would using this model represent extrapolation?

(1) $x = 26$ (3) $x = 14$

(2) $x = 50$ (4) $x = 6$

____

4. Which of the following is true about the line of best fit for the data set given in roster form below?

(1) It has a positive slope and negative y-intercept.
(2) It has both a positive slope and y-intercept.
(3) It has both a negative slope and y-intercept.
(4) It has a negative slope and positive y-intercept.

$$\{(0,-3),(2,4),(6,10),(15,12)\}$$

____

APPLICATIONS

5. An agronomist is studying the height of a corn plants as a function of the number of days since the corn germinated (appeared above the ground). Based on the following data, use your calculator to determine the best fit line in $y = ax + b$ form. Round all coefficients to the nearest *tenth*.

Time, x (days)	3	8	12	20	28	32	40
Height, y (inches)	2.5	4.5	6.2	9.3	12.9	14.4	16.8

6. Heavier cars typically get worse gas mileage (their miles per gallon) than lighter cars. The table below gives the weight versus the highway gas mileage for seven vehicles.

Vehicle Weight (thousands of pounds)	2.5	2.9	3.1	3.0	4.2	6.6	3.4
Gas Mileage (miles per gallon)	34	36	31	29	23	12	26

(a) Determine the best fit linear equation, in $y = ax + b$ form, for this data set. Round all coefficients to the nearest tenth.

(b) Using your model from part (a), determine the gas mileage, to the nearest mile per gallon, for a vehicle that weighs 3500 pounds.

(c) Is the prediction you made in (b) and example of interpolation or extrapolation? Explain.

(d) What is the value of the correlation coefficient to the nearest *hundredth*? Why is it negative?

7. The superintendent of the Clarksville Central School District is attempting to predict the growth in student population in the coming years. The table below gives the population for her district for selected years.

Year	1990	1992	1995	1997	2002	2005
District Population	3520	3605	3771	3860	4135	4285

(a) Find the equation for the line of best fit, in $y = ax + b$ form, where x represents the years *since* 1990 and y represents the district's population. Round all coefficients to the nearest *hundredth*.

(b) Use your model from part (a) to predict the district's population in the year 2020. Round your answer to the nearest whole number.

(c) What are the units of the slope of this linear model?

(d) What does the slope of this model represent? Think about your answer to part (c).

Name: ___________________________________ Date: _________________

OTHER TYPES OF REGRESSION
COMMON CORE ALGEBRA II

Just as we fit data with a linear model we can also fit with all sorts of other mathematical models, depending on the context of the situation. In this lesson we will examine **exponential regression** and **sinusoidal regression**. Exponential regression is review from Common Core Algebra I, so we will start with that.

***Exercise* #1:** The population of Jamestown has been recorded for selected years since 2000. The table below gives these populations.

Year	2002	2004	2005	2007	2009
Population	5564	6121	6300	6812	7422

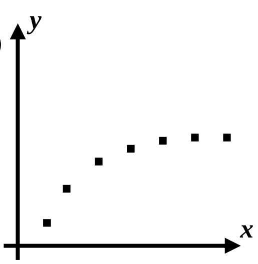
(a) Using your calculator, determine a best fit exponential equation, of the form $y = a \cdot b^x$, where x represents the number of years since 2000 and y represents the population. Round a to the nearest integer and b to the nearest *thousandth*.

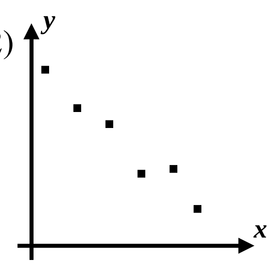
(b) Sketch a graph of the exponential function for the years 2000 to 2050. Label your window and your y-intercept.

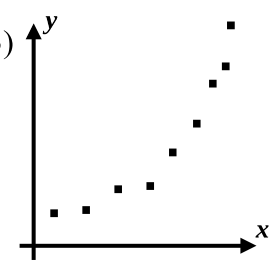
(c) By what percent does your exponential model predict the population is increasing per year? Explain.

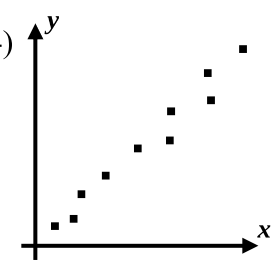
(d) Algebraically determine the number of years, to the nearest year, for the population to reach 20 thousand.

***Exercise* #2:** Which of the following scatter plots would be best fit with an exponential equation?

(1)


(2)


(3)


(4)


Sinusoidal, or trigonometric, regression is much more complicated than either linear or exponential. It should be used in situations that appear **periodic** in nature.

***Exercise* #3:** The temperature of a chemical reaction changes during the reaction. The temperature was measured every two minutes and the data is shown in the table below.

Time (min)	0	2	4	6	8	10	12	14	16	18	20
Temp $(^\circ C)$	35.7	38.9	41.6	42.3	40.8	38.4	36.1	34.2	35.9	39.1	41

(a) Why does it seem like this data might be periodic? Create a quick scatter plot using your calculator to verify.

(b) Use your calculator to do a sine regression in the form $y = a\sin(bx + c) + d$. Round all parameters to the nearest tenth. Graph along with your data to informally assess the fit of the curve.

(c) According to this model, what is the range in temperatures the chemical reaction will include?

(d) According to this model, what is the time it takes for the reaction to complete one full cycle?

Graphing calculators vary. Many will require that if the period of the sinusoidal function is **unknown** then the data must have inputs that are separated by equal amounts (equal steps between *x*-values). On the other hand, there are many periodic phenomena that we want to model whose periods are known. In this case, we can enter data at irregular input intervals.

***Exercise* #4:** The maximum amount of daylight that hits a spot on Earth is a function of the day of the year. Taking $x = 0$ to be January 1st, daylight, in hours, was measured for 12 different days. The measurement was the number of possible hours of sun from sunrise to sunset.

Day	0	34	68	98	118	134	171	203	274	321	346
Daylight Hours	9.0	9.9	11.5	13.1	14.0	14.6	15.2	14.8	13.1	11.5	9.5

(a) What is the natural period of this data set?

(b) Use your calculator with the period from (a) to find an equation of the form $y = a\sin(bx + c) + d$ that fits this data, then examine the graph of the equation on the scatter plot. How good is the fit?

(c) What is the maximum amount of daylight hours predicted by the model? Show your calculation.

Name: ___________________________________ Date: _________________

OTHER TYPES OF CORRELATION
COMMON CORE ALGEBRA II HOMEWORK

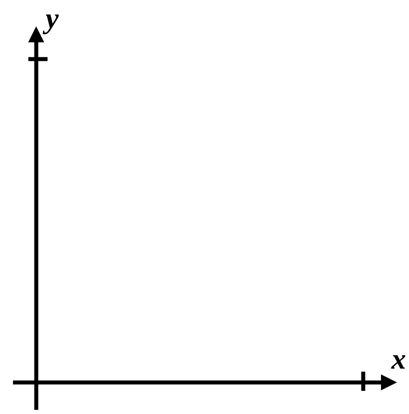
APPLICATIONS

1. Rabbits were accidently introduced to an island where their population is growing rapidly. Biologists studying the rabbits have periodically recorded their population since they were introduced to the island. The data they took is shown below.

Years Since Introduction, x	2	5	7	11	15
Population of Rabbits, y	75	100	112	205	290

(a) Determine an exponential regression equation, in the form $y = a \cdot b^x$, that models this data. Round a to the *tenth* and b to the *hundredth*.

(b) Sketch a graph of the rabbit population below on the axes provided for $0 \le x \le 20$. Label your graphing window and your y-intercept.



(c) Based on your model in part (a), by what percent is the rabbit population growing each year?

(d) Graphically determine, to the nearest *tenth* of a year, when the rabbit population will reach 350.

2. The infiltration rate of a soil is the number of inches or water per hour it can absorb. Hydrologists studied one particular soil and found its infiltration rate decreases exponentially as a rainfall continues.

Time, t (hours)	0	1.5	3.0	4.5	6.0
Infiltration Rate, I (inches per hour)	5.3	3.1	2.4	1.6	0.7

Create an exponential model that best fits this data set. Round parameters to the nearest *hundredth*. Use your model to algebraically determine the time until the rate reaches 0.25 inches per hour. Round your answer to the nearest *tenth* of an hour. Use a logarithm in the process of your algebraic solution.

3. The soil's temperature beneath the ground varies in a periodic manner. A temperature probe was left 3 feet underground and recorded the temperature as a function of the number of days since January 1st ($x = 0$). The temperatures for 14 days throughout the year are shown below.

Day	5	36	57	94	127	153	192
Temp ($^\circ$F)	41	37	36	40	48	64	68
Day	226	241	262	289	305	337	356
Temp ($^\circ$F)	66	61	58	49	44	42	40

(a) Find a best fit sinusoidal function for this data set in the form $y = a\sin(bx + c) + d$. Round all parameters to the nearest *hundredth*. Recall that some calculators require that you input the period on this correlation (365 days).

(b) Based on your model from (a) what are the highest and lowest temperature reached in the soil?

(c) What is the average soil temperature?

(d) If the root of a particular plant species will only thrive when the soil temperature is above 50 $^\circ$F, graphically determine the interval of days over which the plant will thrive.

4. The rise and fall of the tides at a beach is recorded at regular intervals. Their period is almost 24 hours, but not exactly. The depth of a tidal marsh was measured over 3-hour time interval and the data is shown below.

Hours (since midnight)	0	3	6	9	12	15	18	21	24
Depth (ft)	5.5	8.0	10.5	11.7	10.8	8.4	5.8	4.3	4.9

Find a sinusoidal model for this data using your calculator. Place it in $y = a\sin(bx + c) + d$ form. Round all coefficients to the nearest *thousandth* (3 decimal places).

According to your model, what is the period of the tides in hours? Recall that $b \cdot P = 2\pi$.

Name: ____________________________________ Date: ___________________

UNIT #13 – STATISTICS
COMMON CORE ALGEBRA II

Part I Questions

1. Subjects in a sleep study are separated by age into two groups. This separation is done to observe which of the following types of variability?

(1) Measurement (3) Induced

(2) Natural (4) Sample __________

2. In which of the following cases would a survey be more appropriate than an experimental study?

(1) A study to determine how age affects the number of hours of sleep a person gets.

(2) A study to determine the most popular show on television on a given night.

(3) A study to determine the commute time to work based on geographic location.

(4) A study to determine if a particular drug lowers cholesterol. __________

3. On the first day of a small local fair, 55 children, 20 adults, and 25 senior citizens were admitted. If children's tickets cost \$5.00 each, adult tickets cost \$8.00 each and senior citizen tickets cost \$6.00 each, what was the mean ticket price for all 100 people who entered?

(1) \$6.35 (3) \$5.85

(2) \$5.20 (4) \$6.33 __________

4. In which of the following situations is a survey least likely to contain bias?

(1) surveying a sample of people leaving a concert about their favorite musicians

(2) surveying the members of a basketball team to determine the average height of high school boys

(3) surveying people leaving a grocery store about their political party affiliation

(4) surveying teenagers who use social networking websites about their favorite communication methods __________

5. In a survey of 236 freshmen, it was found that 151 of them owned cell phones. Which of the following is closest to the proportion of freshmen who do not own cell phones?

(1) 0.21 (3) 0.43

(2) 0.36 (4) 0.64 __________

6. Students did poorly on a recent test, so their teacher decided to add 6 points to each student's grade. Which of the following statistical measures would not be affected by the addition of these points?

(1) the mean score
(2) the first quartile
(3) the median score
(4) the standard deviation of the scores

____

7. In 2013, the mean gas mileage for cars was 27.6 miles per gallon. If the distribution of gas mileage in cars is normal with a standard deviation of 3.8 miles per gallon, then what percent of cars had gas mileages between 20 and 30 miles per gallon?

(1) 28%
(2) 56%
(3) 71%
(4) 98%

____

8. The gestation time (number of days before birth) for cows is normally distributed with a mean of 284 days and a standard deviation of 12 days. At a local ranch, over the course of a year there are 820 calf births. Of these, how many would be expected to have a gestation time less than 270 days?

(1) 12
(2) 78
(3) 100
(4) 237

____

9. A value's percentile rank is the percent of a data set that lies at or below it. On a standardized test where the scores were normally distributed, Jeremy's score was 1.75 standard deviations above the mean. Which of the following is closest to his percentile rank?

(1) 54^{th}
(2) 67^{th}
(3) 83^{rd}
(4) 96^{th}

____

10. If a distribution of sample means was created from a population, the standard deviation of this distribution would be

(1) equal to the standard deviation of the population
(2) smaller than the standard deviation of the population
(3) larger than the standard deviation of the population
(4) equal to the standard deviation of the population divided by the size of the sample

____

Free Response Questions

11. Mr. Richmond's traffic engineering class is trying to determine people's attitudes towards their evening commute. Students in his class decide to stop drivers on their way home to conduct this survey. Why would this survey method introduce bias into their results?

12. At a local PTA meeting, a sample of parents were surveyed to determine how many children they currently had attending school. Their results are shown in the frequency table below:

Determine the mean, median, and standard deviation for this sample. Round any non-integer answers to the nearest tenth.

Number of Children	Number of Families
1	16
2	24
3	8
4	3
5	2
7	2

Determine how many of the 55 families surveyed have a number of children that was within one standard deviation of the mean. Show your analysis.

13. The scores on a standardized test that Jeremy took were normally distributed with a mean of 82 and a standard deviation of 5. On the test, Jeremy scored a 90.

(a) What percent of students scored better than Jeremy on this test? Round to the nearest tenth of a percent.

(b) If Lisa took the same test, at a different time, and the scores were again normally distributed with a mean now of 83 and a standard deviation of 6.4, then what score, to the nearest integer, would make her percentile rank the same as Jeremy's? Show how you arrived at your answer.

14. Environmental engineers are trying to determine the characteristic fuel economy of cars on the road today. They surveyed 250 drivers about their cars and found the following distribution of fuel efficiencies as rated by the miles per gallon that a given car used while driving on the highway.

Fuel Efficiency (mpg)	Number of Cars
12	2
16	5
18	20
19	35
22	68
26	52
29	30
32	18
45	5

Find the mean and standard deviation for this sample of cars. Round both answers to the nearest *hundredth* of a mile per gallon.

Determine the percent of these cars that fall within one standard deviation of the mean.

Would this sample be well modeled by a normal distribution? Explain your response.

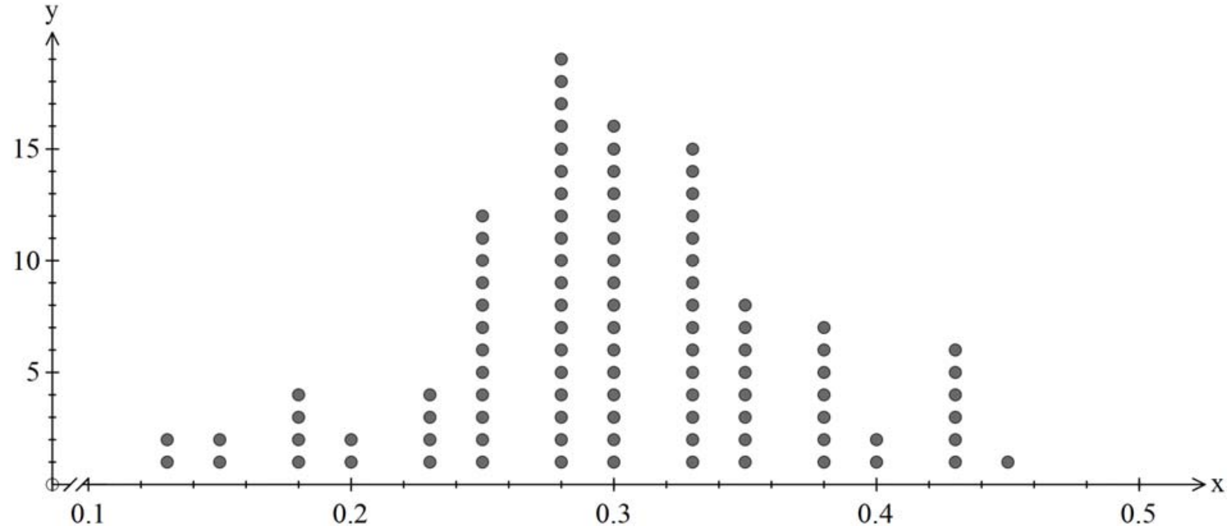
15. Water is flowing out of a reservoir such that the depth of the water is a decreasing function of the number of hours since water was released. Engineers measure the depth of the water and their results are shown in the table below.

Time, x (hours)	2	4	8	14	20
Depth of Water, y (ft)	44.7	36.8	29.2	22.3	15.1

(a) Find an exponential equation, of the form $y = a(b)^x$, that best fits this data set. Round your coefficients to the nearest *hundredth*. Then, use your equation to predict the depth of water after 2 days have elapsed. Round your depth to the nearest *tenth* of a foot.

(b) How would you characterize the strength of the exponential correlation between the two variables? Explain.

16. A survey of 40 high school students was done to determine how many of them liked fresh fruit for lunch. The school will offer a fresh fruit option if more than 30% of students like fruit. Of the 40 surveyed, only 9 of them stated that they liked fruit with lunch. Simulations were done with a population proportion of 0.3 and a sample size of 40 to see how likely a sample of 40 would have only 9 who liked fruit. The results of 100 simulations are shown below.



Simulated Sample Proportions

(a) What was the observed sample proportion for this survey? Round to the nearest hundredth.

(b) If the true population proportion is 0.30, then how likely is it, based on this simulation, that a sample of size 40 would have 9 or fewer students say they like fruit for lunch?

(c) Based on this survey, should the school conclude that they should not serve fruit for lunch? Explain your reasoning.

17. A population that is normally distributed has a mean of 164 and standard deviation of 18.65. If a sample of size 50 was taken from this population, what is the probability its mean would be greater than 168? Show how you arrived at your answer. Round to the nearest tenth of a percent.

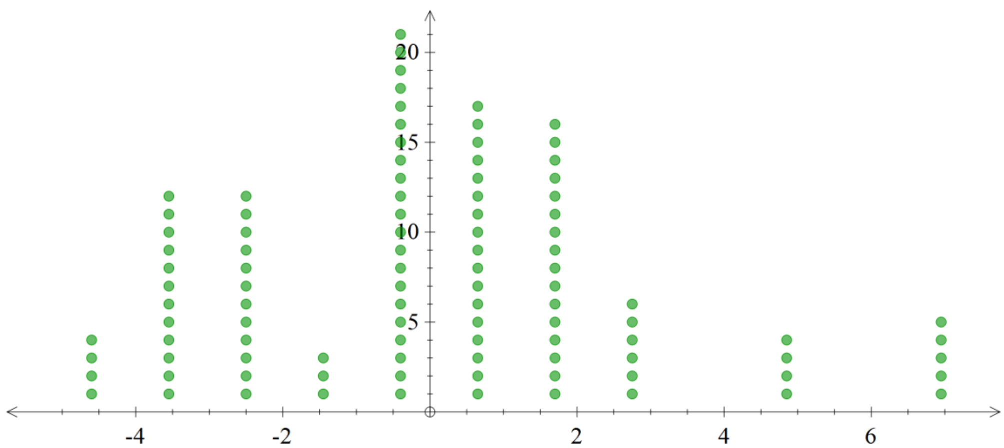
18. In a recent experimental study, 40 participants were broken into two groups. Group 1 maintained a regular exercise routine while Group 2 maintained the same exercise regime but also drank a cup of green tea each day. At the end of a month, the weight loss of each of the 40 subjects was recorded and means for each group were produced as follows:

Group 1: $\overline{x_1} = 5.3$ pounds

Group 2: $\overline{x_2} = 8.6$ pounds

(a) What was the value of $\overline{x_2} - \overline{x_1}$? Using proper units, explain the meaning of this calculation in terms of this experimental study.

(b) The weight losses from the 40 participants were randomly shuffled into two treatment groups 100 times and the distribution of the differences in sample means, specifically $\overline{x_2} - \overline{x_1}$, is shown below.



Difference in Shuffled Treatment Means

Based on this distribution, how significant is the induced (treatment) variability to the weight loss of the participants compared to natural variability? Explain your thinking.

19. In a recent survey of 500 people, 45% of them reported that they were going to vote for Candidate A in an upcoming election. If the survey was reported to have a margin of error of 2%, explain what this means in terms of the actual support for Candidate A.

Name: ___________________________________

Score: ________ / **20**

UNIT #13 QUIZ (THROUGH LESSON #6)
COMMON CORE ALGEBRA II

1. Mrs. Rose randomly selects 15 equal sized groups of students from a list of all of her current students. For each group, she looks up their scores on the Geometry Regents and then calculates each group's mean score. Which of the following types of variability would most likely explain any differences in these mean scores?

(1) natural (3) sampling

(2) induced (4) measurement

____

2. In a survey of 214 juniors, it was found that 139 of them are taking Algebra II. Which of the following is closest to the proportion of juniors who are *not* taking Algebra II?

(1) 0.35 (3) 0.54

(2) 0.46 (4) 0.65

____

3. A data set is normally distributed with a mean of 26 and a standard deviation of 4. Which of the following is closest to the percent of the data set that is between 22 and 28?

(1) 34.1% (3) 53.2%

(2) 38.2% (4) 68.2%

____

4. The ages of students in a club are normally distributed with a mean of 14 years and a standard deviation of 2.6 years. Which of the following is closest to the z-score for a student who is 17 years old?

(1) 0.50 (3) 0.87

(2) 0.67 (4) 1.15

____

5. Which of the following best describes the mean of the sample means after running a large number of simulations?

(1) It is equal to the population mean.

(2) It is greater than the population mean.

(3) It is less than the population mean.

(4) It may be greater or less than the population mean.

____

6. Lydia and Shannan wanted to explore whether people are more productive while listening to classical music or while working in complete silence. Twenty people were randomly split into two groups and were asked to work on a math puzzle. One group was given noise-cancelling headphones while the other group listened to classical music. After one hour, they collected the puzzles and analyzed the results. Shannan says that they performed an observational study, but Lydia disagrees and says that it was an experimental study. Which girl is correct? Explain your reasoning. [2 points]

7. A group of 50 students took a college placement test to determine which math course they should take during their freshman year. The results are displayed in the frequency table below.

Score	Frequency
60	5
65	10
70	8
75	6
80	6
85	8
90	3
95	2
100	2

(a) Find the mean and the population standard deviation for this set of test scores. Round to the nearest tenth. [2 points]

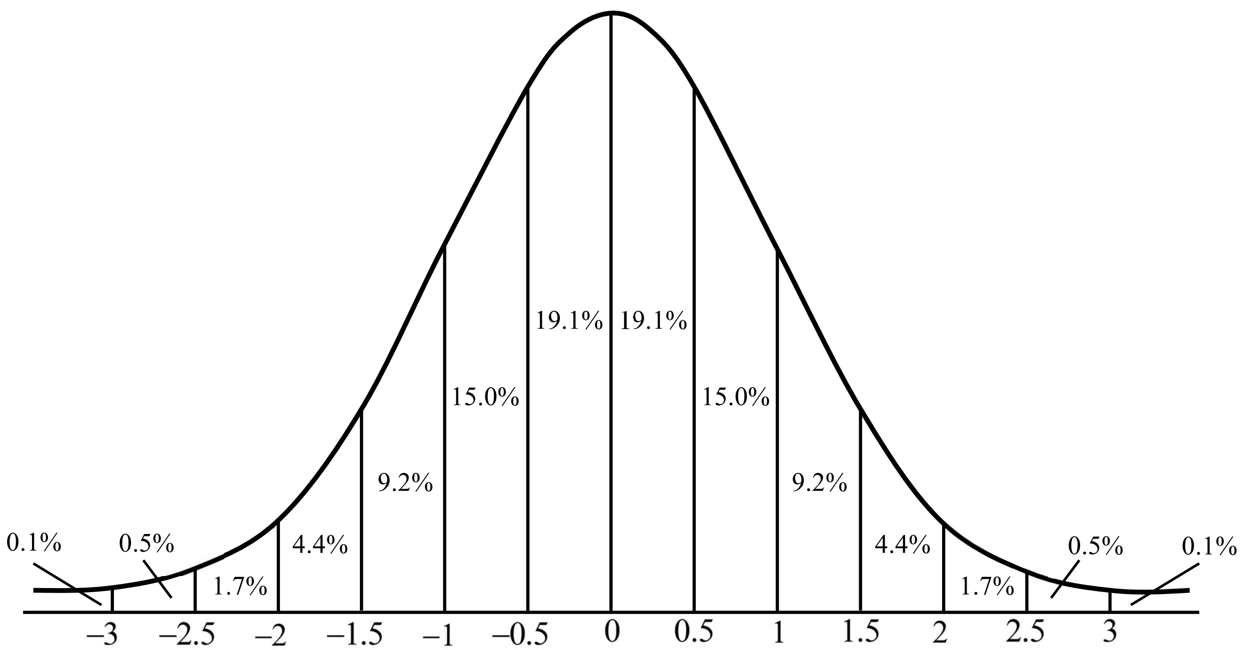
(b) Determine how many scores fall within one standard deviation of the mean. [2 points]

(c) Determine the interquartile range for this set of test scores. [2 points]

8. At a high school, it is known that 76% of all students participate in at least one club. A survey was randomly given to 42 freshmen in their study hall, and 31 students stated that they participate in at least one club. Which is higher, the population proportion of students participating in at least one club or the sample proportion of freshmen who participate in at least one club? Justify your response. [2 points]

THE NORMAL DISTRIBUTION

BASED ON STANDARD DEVIATION



Name: ___________________________________

Score: ________ / 30

UNIT #13 ASSESSMENT
COMMON CORE ALGEBRA II

Part I Questions

1. On a recent Algebra II test, the mean score was a 74. Mr. Weiler decides to add 6 points to each student's score. Which of the following statistical measures will not change after the addition of the 6 points?

   (1) the mean score (3) the median score

   (2) the first quartile (4) the standard deviation

   ____

2. A group of 40 subjects was divided into two groups with one group given an energy drink and one group given water. The subjects were then given a spelling test and the mean number of words correctly spelled was recorded for each member of the two groups. This experiment was designed this way to quantify which of the following types of variability?

   (1) natural (3) induced

   (2) sampling (4) measurement

   ____

3. The length of pop music songs is normally distributed with mean length of 215 seconds and a standard deviation of 38 seconds. What percent of pop songs have lengths between three and four minutes?

   (1) 48% (3) 61%

   (2) 57% (4) 67%

   ____

4. In which of the following scenarios would an observational study be used instead of a survey or experiment?

   (1) a study to quantify the effect of humidity levels on plant growth

   (2) a study to determine how age effects political preference

   (3) a study to measure the level of support for a ballot referendum

   (4) a study to observe the effects of a medicine designed to lower cholesterol __________

5. A population of emperor penguins has a mean weight of 77 pounds with a standard deviation of 16 pounds. What would be the standard deviation of sample means taken from this population with a sample size of 50?

   (1) 0.32 (3) 2.26

   (2) 1.54 (4) 10.89

   ____

6. If an exponential model was used to fit the data set below, which of the following would be the best prediction for the output of the model if the input was $x = 20$?

(1) 766

(2) 1,164

(3) 1,276

(4) 1, 488

x	3	7	11	14	17
y	83	142	301	450	722

____

7. In a recent poll of 350 likely voters, 42% of them preferred the incumbent candidate. Which of the following would be closest to the margin of error of this statistic?

(1) 2.6%

(2) 3.7%

(3) 4.2%

(4) 5.3%

____

8. A population of fruit flies has a mean life span of 46 days with a standard deviation of 6.2 days. If a sample of 30 fruit flies is taken from this population, what is the probability it will have a sample mean life span of greater than 48 days?

(1) 3.9%

(2) 5.3%

(3) 28.7%

(4) 36.2%

____

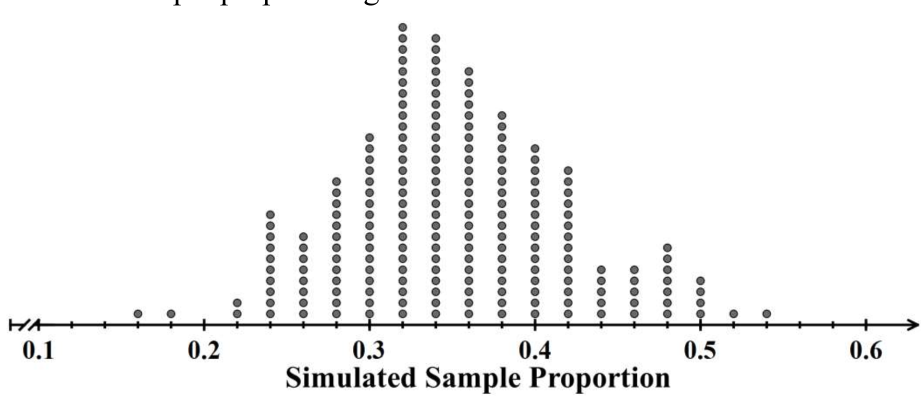
9. A simulation was run 200 times using a population with a proportion of 0.35 and a simulated sample size of 50. The simulated sample proportions are shown below. Based on this distribution, what is the probability of obtaining a sample of 50 with a sample proportion greater than 0.45?

(1) 4.5%

(2) 9%

(3) 14%

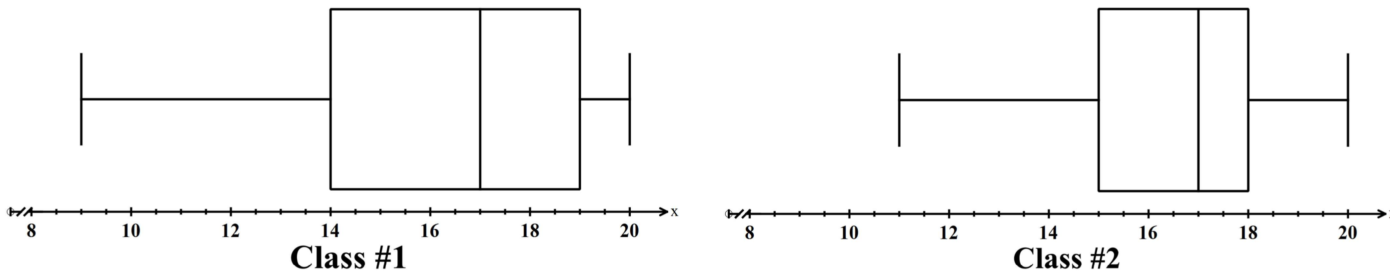
(4) 18%



____

PART II QUESTIONS: Answer all questions in this part. Each correct answer will receive 2 credits. Clearly indicate the necessary steps and explain your reasoning. For all questions in this part, a correct numerical answer with no work shown will receive only 1 credit.

10. A 20 point test was given to two classes. Each class had results shown in the box plots below. Which of the two classes has the greater interquartile range?



11. On a standardized test where the results are distributed normally, Elliete scored a 246. If the mean score was a 182 with a standard deviation of 36, then what is Elliete's percentile rank rounded to the nearest whole percent? Explain how you arrived at your answer.

12. A sample of 38 students is chosen from a student body where 58% of all students own cell phones. If 24 of the students in this sample own cell phones, which is higher, the population proportion or the sample proportion?

13. What is the amplitude of the best fit sinusoidal model for the data set shown below? Round your answer to the nearest *hundredth*.

x	2	5	8	11	14	17
y	10	3	7	9	2	8

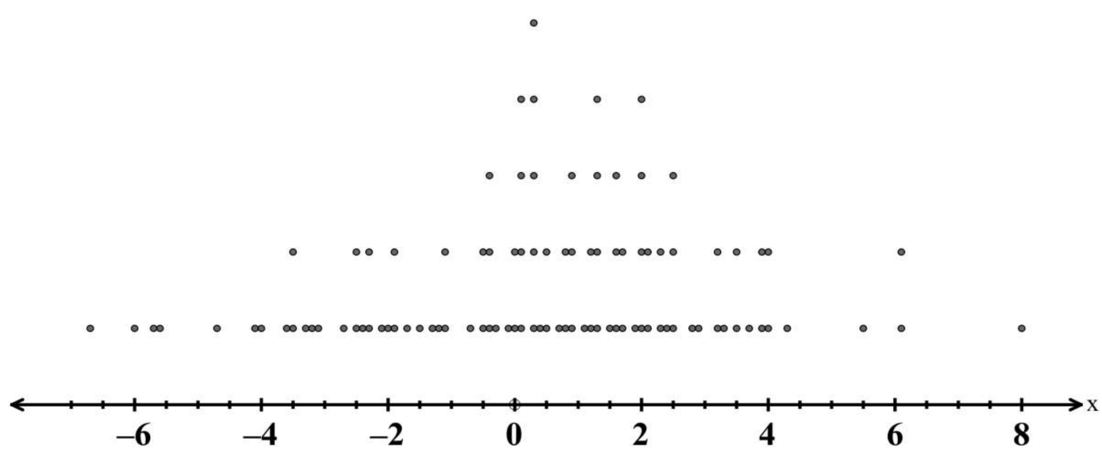
PART III QUESTIONS: Answer all questions in this part. Each correct answer will receive 4 credits. Clearly indicate the necessary steps and explain your reasoning. For all questions in this part, a correct numerical answer with no work shown will receive only 1 credit.

14. Mia believes that corn plants grown with synthetic light will have a slower growth rate than those grown in natural light. She randomly plants 30 seeds, 15 with synthetic light and 15 with natural light. After 4 weeks of growth, she measures each plant's height in centimeters and finds the following:

Treatment #1 - Synthetic Light	Treatment #2 - Natural Light
$\overline{x_1} = 17.9$ cm	$\overline{x_2} = 23.4$ cm
$s_1 = 7.9$ cm	$s_2 = 5.2$ cm

(a) Calculate $\overline{x_2} - \overline{x_1}$. Does this calculation support Mia's belief? Explain. [2 points]

(b) Mia randomly shuffles the 30 results into two groups and repeatedly calculates $\overline{x_2} - \overline{x_1}$ for each simulation. She does this 100 times with the results shown below.



Does the simulation and the results from (a) show a statistically significant increase in plant growth due to the natural light treatment? Explain.