Math 3 Unit 5: Sampling & Data Analysis

Unit	Title	Standards
5.1	Sampling & Study Methods	S.IC.1, S.IC.3
5.2	Standard Deviation	S.1D.1, S.ID.2, S.1D.3
5.3	Modified Box and Whisker Plots	S.1D.1, S.ID.2, S.1D.3
5.4	Normal Distribution, Empirical Rule (68-95-99.7 Rule) & Z-Scores	S.ID.4
Unit 5 Review		

Math 3 Unit 5B: Regression

Unit	Title	Standards
5B.1	Linear Regression & Line of Best Fit	S.ID.6a
5B.2	Other Types of Regression	S.ID.6c
5B.3	Quantifying Predictability	S.ID.6, S.ID.8

Additional Clovis Unified Resources
http://mathhelp.cusd.com/courses/math-3

Clovis Unified is dedicated to helping you be successful in Math 3. On the website above you will find videos from Clovis Unified teachers on lessons, homework, and reviews. Digital copies of the worksheets, as well as hyperlinks to the videos listed on the back are also available at this site.

Math 3 Unit 5: Online Resources

5.1	Sampling & Study Methods	• Statistics Learning Centre: Sampling: Simple Random, Convenience, Systematic, Cluster, Stratified - Statistics Help http://bit.ly/51sasma • Khan Academy - Article: Sampling Methods Review http://bit.ly/51sasmb • Khan Academy - Article: Reasonable Samples http://bit.ly/51sasmc • Changing Minds: Choosing a Sampling Method http://bit.ly/51sasmd
5.2	Standard Deviation	• Patrick JMT: Statistics: Calculating Variance http://bit.ly/52stdva • Khan Academy: Sample Standard Deviation and Bias http://bit.ly/52stdvb • James Soares: Find the Standard Deviation http://bit.ly/52stdvc • Math is Fun: Standard Deviation & Variance http://bit.ly/52stdvd
5.3	Modified Box and Whisker Plots	• Patrick JMT: Box and Whisker Plot http://bit.ly/53mbwpa • Khan Academy: Constructing a Box Plot & Interpreting Box Plots http://bit.ly/53mbwpb & http://bit.ly/53mbwpc • Purple Math: Box-and-Whisker Plots: Interquartile Ranges and Outliers (Page 3) http://bit.ly/53mbwpd • James Soares: Create a Modified Box and Whisker Plot & Find Percentiles http://bit.ly/53mbwpe & http://bit.ly/53mbwpf • Purple Math: Box-and-Whisker Plots: Five-Number Summary (Pages 1-2) http://bit.ly/53mbwpg • James Soares: Find the 5 Number Summary http://bit.ly/53mbwph
5.4	Normal Distribution, Empirical Rule (68-95-99.7 Rule) & Z-Scores	• Patrick JMT: The Normal Distribution and the 68-95-99.7 Rule http://bit.ly/54neza • James Soares: The Normal Distribution http://bit.ly/54nezb • Khan Academy: F Normal Distribution Problems: Empirical Rule (from ck12.org) http://bit.ly/54nezc • Khan Academy: Normal Distribution Problem: Z-scores (from ck12.org) http://bit.ly/54nezd • Khan Academy - Article: Z-Scores Review http://bit.ly/54neze
5B.1	Linear Regression & Line of Best Fit	• eMathInstuction: Common Core Algebra II - Linear Regression and Lines of Best Fit http://bit.ly/5b1lra • Purple Math: Scatterplots and Regressions http://bit.ly/5b1lrb • Khan Academy: Fitting a Line to Data http://bit.ly/5b1lrc
5B.2	Other Types of Regression	• eMathInstuction: Common Core Algebra I - Other Types of Regression http://bit.ly/5b2ora • Khan Academy: Comparing Models to Fit Data Example http://bit.ly/5b2orc
5B.3	Quantifying Predictability	• eMathInstuction: Common Core Algebra I - Correlation Coefficients http://bit.ly/5b3qpa • Khan Academy: Example: Correlation Coefficient Intuition & Correlation Coefficient Review http://bit.ly/5b3qpb & http://bit.ly/5b3qpc

[1- 6] Identify the type of **sampling method** used in each example below.   Then identify any bias in each method.

1.   The manager of Save Mart wants to find the percent of shoppers who bring their own bags. He interviews every shopper entering the produce section.

2.   A school principal gathers an alphabetical list of all the students at her school. Then she selects every 15th student to take a survey about the cafeteria's lunch menu.

3.   An online advertisement asks you to participate in a survey. The survey asks how much time you spend online each week.

4.   A magazine publisher mails a survey to every sixth person on a subscriber list. The survey asks for four favorite leisure activities.

5.   Student leadership members survey every tenth student who enters the school building and ask whether students favor the new drew code.

6.   To determine a community's reading habits, a newspaper conducts a poll from a table near the exit of a history museum.

7.   Describe how a convenience sample and a self-selected sample are alike and how they are different.

[8-10]. Identify the type of **study method** described in each situation and explain whether the sample statistics could be used to make a general conclusion about the population.

8.   A school psychologist sits in a school cafeteria and takes notes on students' behavior while they eat lunch.

9.   A researcher places 50 houseplants of the same type and size in each of 2 identical soundproof rooms. Each plant receives the exact same amount of water and light. Classical music is played continuously in one room, and no music is played in the other room. After 30 days, the researcher measures the growth of the plants.

10.   A company selects every 20th name from an alphabetical list of the company's employees. The employees selected are asked whether they are satisfied with the company's benefits package.

[11-14] Answer the following:

11.   A survey asks, "Aren't healthy, organic vegetables always better than non-organic vegetables?" Does this survey question have any bias? Explain.

12.   What is the difference between a sample and a population? Give an example of each.

13.   What does it mean to have an unbiased sample? Why does it matter?

14.   Would a large or small sample tend to give a better estimate of how the total population feels about a topic? Explain.

[15-18] Creating Sampling Methods

Below are some hypothetical situations. Describe in detail how to obtain a sample. Choose one sampling method for each example below. Use each sampling method once and only once. Determine if the sample obtained will be biased or unbiased. If there is bias, explain why.

Convenience   **Self-selected**   **Systematic**   **Random**

15. You would like to survey the students at your high school and find out what percent of the students have an iPhone.

16. A grocery store would like to find out the average amount of time a customer spends in line before he/she pays for the products purchased.

17. GoPro makes a small digital video camera that is used quite a bit around the world. You are a quality control inspector for the company. You are wondering about the reliability of the batteries made for the cameras.

18. You are at a town hall meeting. The meeting has approximately 1,300 people attending regarding water usage during the summer months. You are responsible for investigating the number of minutes used in a day to water their lawn and plants around their homes. You would like to, quickly, gather a sample of residents and ask them questions after the meeting is done.

[19-21] Answer the following:

19. In the last election, a survey of randomly selected registered voters was conducted to determine which candidate was likely to win. 49% of respondents chose candidate A and 45% chose candidate B with 6% undecided. Based on the sampling results, can you make a general conclusion that it is more likely that candidate A will win the election? Explain.

20. Best Buy conducts a survey of customer satisfaction. 400 customers are randomly selected from a list of 3000 new customers in the past year. Of the 400 surveys sent, 200 hundred are returned. The statistics showed a high level of customer satisfaction. Can Best Buy assume this is true overall? Explain.

21. A student wants to determine the percentage of Clovis Unified students who drive a car to school. She observes how many students drive into the parking lot of her school before school on Monday. What type of sampling method was used? What type of study was performed? Can the student use these statistics to make a general conclusion about all the students in Clovis Unified? Explain.

Math 3 Unit 5 Worksheet 2
Standard Deviation

Name:_____
Date:_____Per:_____

1. Two students, Elvira and Bob, were in the same Psychology class at Clearwater State University. There were ten quizzes taken during the semester. The students' scores were as follows:

Elvira's Scores:

85	87	90	84	87	85	84	86	87	85

Bob's scores:

53	97	75	80	90	45	71	99	60	90

a) Which student do you think had the lower standard deviation?_____

b) Justify your answer:_____

_____

c) If the average for Elvira's exam scores is 86, find the Standard Deviation for Elvira's exam scores. (Round your answers to the nearest thousandth.)

Scores x	Average μ	$x - \mu$	$(x - \mu)^2$
85	86		
87	86		
90	86		
84	86		
87	86		
85	86		
84	86		
86	86		
87	86		
85	86		
			$\Sigma(x - \mu)^2 =$

Variance(σ^2):_____
$\frac{\Sigma(x-\mu)^2}{n}$

Standard Deviation(σ):_____

d) Find Bob's mean, variance and standard deviation: (Round your answers to the nearest thousandth.)

x	μ	$x - \mu$	$(x - \mu)^2$
			$\Sigma(x - \mu)^2 =$

Variance(σ^2):_____
$\frac{\Sigma(x-\mu)^2}{n}$

Standard Deviation(σ):_____

e) How many quiz scores for Bob were within one standard deviation of the mean?

f) Summarize what you discover when you compare the standard deviation of the two students quiz scores.

g) Summarize what you discover when you compare the ranges of the two students quiz scores.

2. a) Find the Mean, Variance, and Standard Deviation from the data given below:
(Round your answers to the nearest hundredth.)

			X			
X			X	X	X	
X	X	X	X	X	X	X
1	2	3	4	5	6	7

x	μ	$x - \mu$	$(x - \mu)^2$
			$\Sigma(x - \mu)^2 =$

Variance(σ^2):_____

$\dfrac{\Sigma(x-\mu)^2}{n}$

Standard Deviation(σ):_____

b) Calculate the NEW mean, NEW Variance, and NEW Standard Deviation if someone revealed the data given below: (Round your answers to the nearest tenth.)

			X			
X			X	X	X	
X	X	X	X	X	X	X
10	20	30	40	50	60	70

x	μ	$x - \mu$	$(x - \mu)^2$
			$\Sigma(x - \mu)^2 =$

Variance(σ^2):_____

$\dfrac{\Sigma(x-\mu)^2}{n}$

Standard Deviation(σ):_____

c) What is the relation between the values for the OLD Mean and Standard Deviation and the NEW Mean and Standard Deviation?

d) Calculate the Mean, Variance, and Standard Deviation from the data given below.
(Round your answers to the nearest hundredth.)

			X			
X			X	X	X	
X	X	X	X	X	X	X
4	5	6	7	8	9	10

x	μ	$x - \mu$	$(x - \mu)^2$
			$\sum(x - \mu)^2 =$

Variance(σ^2):_____

$\dfrac{\sum(x-\mu)^2}{n}$

Standard Deviation(σ):_____

e) Based upon the problems above, complete the table below:

Data Analysis	Multiply each item by a constant	Add each item by a constant
Mean	*Change: Multiply the original mean by the constant.*	
Standard Deviation		

f) What is the standard deviation of a class if everyone in the class scored an 85% on the exam? Why?

3. Find the mean, variance and standard deviation for each data set:
   (Round your answers to the nearest hundredth.)

   a)   13   14   17   18   12   21   10

   b)   12   3   2   4   5   7

   c)   60   40   35   45   39

Math 3 Unit 5 Worksheet 3   Name:_____

Modified Box and Whisker Plots   Date:_____Per:_____

1. The grade point averages for 20 students are listed below.

| 2.0 | 3.2 | 1.8 | 2.9 | 1.9 | 4.0 | 3.3 | 2.9 | 3.6 | 0.8 |
| 3.1 | 2.4 | 2.4 | 2.3 | 1.6 | 1.6 | 4.0 | 3.1 | 3.2 | 1.8 |

a) Use the data above to create a 5-number summary

Min	Q1	Med	Q3	Max

b) Determine the Interquartile Range (IQR), Lower Boundary, and Upper Boundary to determine and identify any outliers.

IQR_____
Lower Boundary_____
Upper Boundary_____
Outlier(s)_____

c) Create a Modified Box and Whisker Plot using the information you determined in parts (a) and (b).
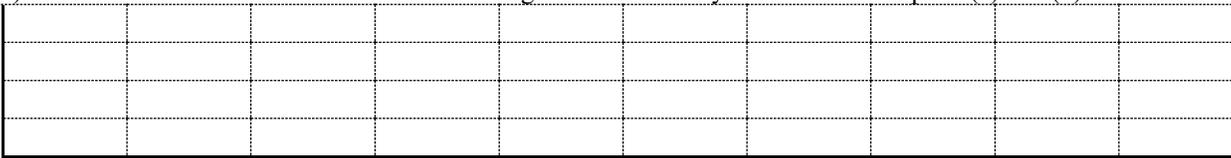
2. The heights of 14 adult males are listed below.

| 70 | 65 | 75 | 69 | 68 | 64 | 69 | 69 | 66 | 71 | 68 | 67 | 73 | 68 |

a) Use the data above to create a 5-number summary

Min	Q1	Med	Q3	Max

b) Determine the Interquartile Range (IQR), Lower Boundary, and Upper Boundary to determine and identify any outliers.

IQR_____
Lower Boundary_____
Upper Boundary_____
Outlier(s)_____

c) Create a Modified Box and Whisker Plot using the information you determined in parts (a) and (b).
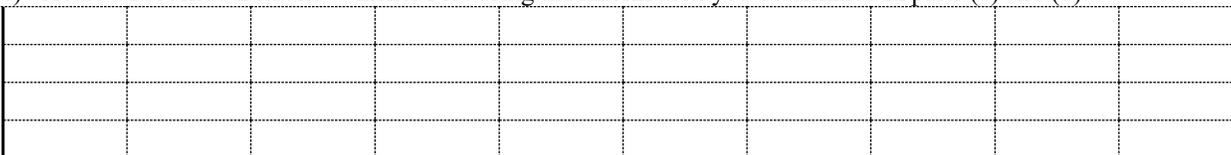
d) If the outlier were removed from the data set, how would this affect the mean of the data set? Explain your answer.

e) If the outlier were removed from the data set, how would this affect the standard deviation of the data set? Explain your answer.

3. The Highway Patrol, using radar, checked the speed (in mph) of 30 passing motorists at a checkpoint.

30	36	36	37	39	40
40	41	41	41	42	42
42	43	43	43	44	45
45	45	47	48	48	48
48	49	49	50	50	50

a) Use the data above to create a 5-number summary

Min	Q1	Med	Q3	Max

b) Determine the Interquartile Range (IQR), Lower Boundary, and Upper Boundary to determine and identify any outliers.

IQR_____
Lower Boundary_____
Upper Boundary_____
Outlier(s)_____

c) Create a Modified Box and Whisker Plot using the information you determined in parts (a) and (b).
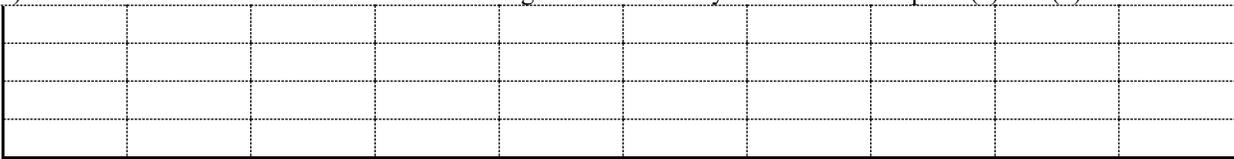
d) If the outlier were removed from the data set, how would this affect the mean of the data set? Explain your answer.

e) If the outlier were removed from the data set, how would this affect the standard deviation of the data set? Explain your answer.

4. Given the following 5-Number Summary for a set of test scores. Let the class size for the data equal 28.

Min	Q1	Med	Q3	Max
53	64	70	75	94

a) Do any outliers exist in the class? Provide mathematical proof of your answer.

b) Does the value 70 need to be an actual data value in the original set of test scores? Explain your answer.

5. Given the following 5-Number Summary for a set of test scores. Let the class size for the data equal 31.

Min	Q1	Med	Q3	Max
57	66	74	83	97

    a) Do any outliers exist in the class? Provide mathematical proof of your answer.

    b) Does the value 74 need to be an actual data value in the original set of test scores? Explain your answer.

6. Suppose The Buck Depot, which currently sells 492 distinct items that cost between $0.25 and $1.00, decides to sell an item that costs $5. Select whether the value of each statistic, for the prices of the items at The Buck Depot, increases, decreases, or cannot be determined.

	Increases	Decreases	Cannot Be Determined	Stays the Same
Mean				
Median				
Mode				
Standard Deviation				

7. Suppose Anne Kersaway, who collects antique sailboats, currently owns 31 sailboats that each have a market value between $13,250 and $127,500.   She then decides to purchase another sailboat that has a market value of $10,135. Select whether the value of each statistic, for the market values of her sailboats, increases, decreases, or cannot be determined.

	Increases	Decreases	Cannot Be Determined	Stays the Same
Mean				
Median				
Mode				
Standard Deviation				

1.   In a normal distribution, what percent of the values lie:

a. below the mean? _____   b. above the mean? _____

c. within one standard deviation of the mean? _____   d. within two standard deviations of the mean? _____

e. within three standard deviations of the mean? _____

2.   1800 students at Cal State University took a biology exam. The scores were distributed normally with a mean of 70 and a standard deviation of 5. Label the mean and three standard deviations from the mean.
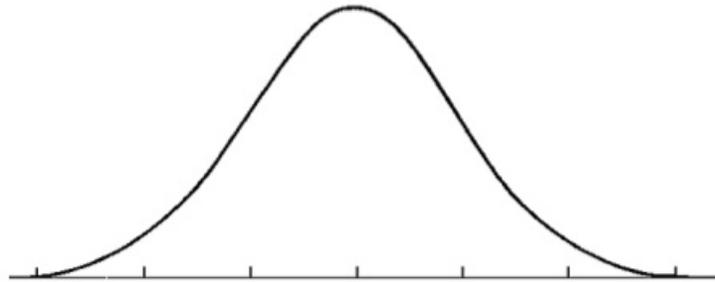
a. What percentage of scores are between 65 and 75?   b. What percentage of scores are between 60 and 70?

c. What percentage of scores are between 60 and 85?   d. What percentage of scores is less than 55?

e. What percentage of scores is greater than 80?

f. Approximately how many biology students scored between 60 and 70?

g. Approximately how many biology students scored between 55 and 60?

h. If Ron Osserus scored 80 on the biology exam, what is his z-score?

i. If Amos Kittow had a z-score of – 1 on this exam, what is his biology exam score?

3.	600 juniors at Union Prep Academy took the ACT last year. The scores were distributed normally with a mean of 22 and a standard deviation of 4. Label the mean and three standard deviations from the mean.
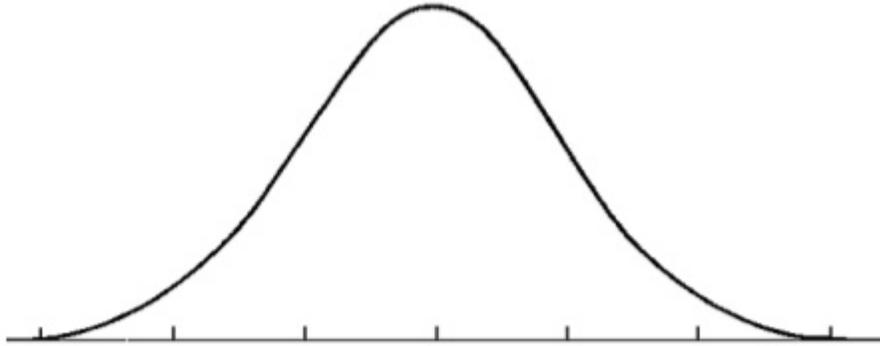
a) What percentage of scores are between scores 18 and 26?

b) What percentage of scores are between scores 14 and 30?

c) What percentage of scores are between scores 14 and 26?

d) What percentage of scores is less than a score of 14?

e) What percentage of scores is greater than a score of 22?

f) Approximately how many juniors scored between 22 and 26?

g) Approximately how many juniors scored between 18 and 26?

h) Approximately how many juniors scored between 22 and 30?

i) Approximately how many juniors scored between 14 and 18?

j) Approximately how many juniors scored lower than 30?

k) If Teerzov Joy had a z-score of 3 on the ACT, what is her actual ACT score?

l) If Beau N. Ero scored 18 on the ACT, what is his z-score?

m) (Without calculator) If Chris P. Bacon scored 27 on the ACT, his z-score would be between the integers ___ and ___ .

n) (With calculator) Now calculate the exact z-score of Chris P. Bacon's score of 27 on the ACT.

4.   Identify the standard deviation and the mean from the given normal distribution.

    0.4   1.1   1.8   2.5   3.2   3.9   4.6
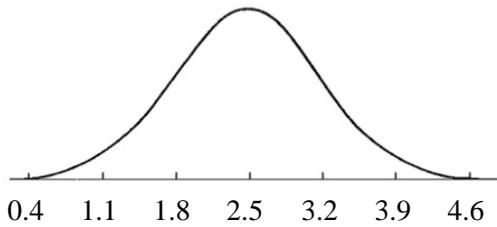
5.   In a normal distribution of test scores, the middle 68% of test takers scored between 66 and 78.
        What is the mean? What is the standard deviation?
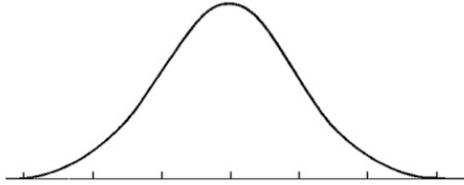
6.   In a normal distribution of Honda Accord prices, the middle 95% of buyers paid between $22,000 and $29,000.
        What is the mean? What is the standard deviation?
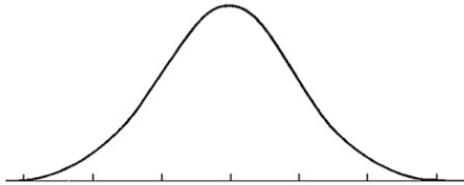
7.   An exam produced grades with a mean score of 74.2 and a standard deviation of 11.5. Find the z-scores, rounded
        to the nearest hundredth, for each of the following test scores. Assume the distribution for the exams to be
        normal. Also, write a sentence interpreting the meaning of each z-score in relation to the mean. (Example
        sentence: "A test score of 60 is 1.23 standard deviations below the mean.")

        a.   test score = 54   b.   test score = 68

        c.   test score = 79   d.   test score = 93

8.   A sample of textbooks has a mean price of $120 and a standard deviation of $20. Find the textbook price that
        corresponds each of the z-scores. Assume the distribution of the textbook prices is normal.

        a. $z = 0.0$ b. $z = 1.2$

        c. $z = -1.4$ d. $z = 2.05$

9.   What does it mean to say that $x = 152$ has a z-score (standard score) of 1.50? Assume the distribution of the
        values to be normal.

10. A sample of restaurants in a Kalfene City revealed that the distribution of the price for a cup of coffee was normal. The price for a cup of coffee at Mug-O-Joe is $3.26 and has a z-value of 2.24. What is the average price for a cup of coffee in the city if the standard deviation was 25 cents?

11. An applicant's score on a company's employment exam was 215. The average score for the exam is 189. If the z-score for this applicant was 2.5, what was the standard deviation for the employment exams? Assume the exam scores for the company to be normal.

12. A married couple are employed by the same company. The husband works in a department for which the mean hourly rate is $12.80 and the standard deviation is $1.20. His wife is employed in a department where the mean rate is $13.50 with a standard deviation of $1.80. Relative to their departments, which is better paid if the husband earns $14.60 and the wife earns 15.75? Explain your answer with mathematical proof/justification. Assume the distribution for both rates is normal.

13. The average yearly rainfall in Los Angeles is 14.95 inches with a standard deviation of 7.02 inches. The average yearly rainfall in Miami is 60.2 inches with a standard deviation of 19.4 inches. The wettest year in history for Los Angeles produced 33.44 inches of rain, while the wettest year in Miami produced 89.33 inches of rain. Relative to each cities weather, which wettest year was less likely to have occurred? Explain your answer with mathematical proof/justification. Assume the rainfall distribution for both cities is normal.

14. For the graph of the data set below, what statistic would be a better measure of the spread of the data, the standard deviation or the interquartile range? Explain your answer.
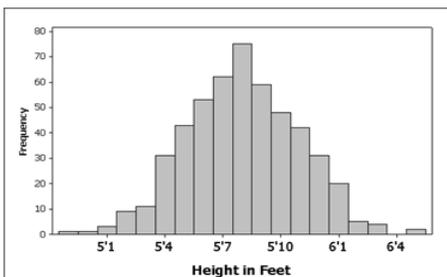
15. For the graph of the data set below, what statistic would be a better measure of the spread of the data, the standard deviation or the interquartile range? Explain your answer.
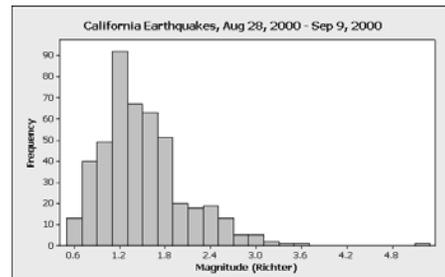
Challenge Question

16. Tim scored a 95% on his exam and had a corresponding z-score of 1.5. Mary scored a 98% on her exam and had a corresponding z-score of 2.0. What is the mean and the standard deviation for the class exam if the distribution is normal?

Math 3 Unit 5 Review Worksheet

Name:_____

Date:_____Per:_____

[1-5] State whether the example is a Convenience, Systematic, Random, or Self-Selected Sample.

1. To find the mean height of all third grade students in a school district, we measure the first five students who are dropped off in the morning by their parents.

2. A student is working on a survey of students' attitudes toward college administrators, and so he talks to his roommate and other people on the floor of his residence hall.

3. A researcher wants to select 10 people from a population of 100. The researcher has a list of all 100 people, the researcher then assigns each person a number from 1 to 100. The researcher then picks a random number, 6, as the starting number. He or she would then select every tenth person for the sample (because the sampling interval = 100/10 = 10). The final sample would contain those individuals who were assigned the following numbers: 6, 16, 26, 36, 46, 56, 66, 76, 86, 96.

4. You have a population of 1,000 people. You wish to choose a sample of 50 people. First, each person is numbered 1 through 1,000. Then, you generate a list of 50 random numbers (with a computer program) and those individuals assigned those numbers are the ones you include in the sample.

5. You ask your parents and friends to take a survey.

[6-12] State whether you should conduct a Survey, Observational Study, or a Controlled Experimental Study.

6. You want to see how much water dogs drink on average per day.

7. The principal wants to know exactly how many students have internet access from their home.

8. The class of 2016 is trying to decide which of three different gifts they are going to present to the school.

9. The City wants to see the effects of lawn being watered one day a week in the summer months compared to the allowed two days a week.

10. A group of students is interested in knowing if the number of times they can sink a basketball is related to the color of the basketball. The students shoot a series of baskets and record their success using a regulation colored basketball. They then switch to a blue colored basketball and shoot the same series of baskets.

11. Some researchers are interested in knowing if exercise prevents colds. They randomly select a sample of people and record the amounts of weekly exercise and numbers of colds last year.

12. A candidate for public office wants to know the percentage of citizens that favor a flat-rate income tax. She sends out a questionnaire to registered voters to gather their opinions.

[13-16] Decide which type of sampling method is being utilized for each study: Random Sampling, Convenience Sampling, Systematic Sampling, or Self-Selected Sampling.

13. A news show wants to get feedback on the popularity of their reporters, so they ask viewers to participate in an on-line poll.

14. Each member of the population is assigned a unique number. The numbers are placed in a bowl and thoroughly mixed. Then, a blind-folded researcher selects 25 numbers from the bowl.

15. To determine the proportion of yellow M&Ms produced by a factory, Kevin goes to the store in his neighborhood and buys the first 10 bags of M&Ms on the shelf. He then counts the M&Ms.

16. A company's Human Resources department wants to pick a sample of employees and ask how they feel about company policies. Starting with the tenth person who arrives to work, every 4th person is selected to fill out a survey.

17. Which of the sampling methods in #s 13-16 are biased? Provide an explanation of the bias.

[18-22] State what the following symbols mean.

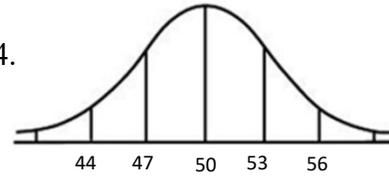18. μ 19. σ 20. $\frac{\Sigma x}{n}$ 21. σ^2 22. $\frac{\Sigma(x-\mu)^2}{n}$

[23-26] Find the standard deviation given the following:

23. | 4, 9, 11, 11, 13, 14, 14, 20 | 24.

44   47   50   53   56

25. $\Sigma(x-\mu)^2 = 84$, and $n = 7$ 26. $\sigma^2 = 16$

27. Find the standard deviation of a data set whose variance is 7.3.

28. Find the standard deviation of a data set if $\Sigma(x-\mu)^2 = 100$ and the data set size is 20.

29. Find the standard deviation of the data set: | 21, 24, 37, 31, 22, 35, 29, 34, 30, 27 |

30.   2 18 19 21   25   27   30   41

a) Use the data above to create a 5-number summary

Min	Q1	Med	Q3	Max

b) Determine the Interquartile Range (IQR), Lower Boundary, and Upper Boundary to determine and identify any outliers.

IQR_____

Lower Boundary_____

Upper Boundary_____

Outlier(s)_____

c) Create a Modified Box and Whisker Plot using the information you determined in parts (a) and (b).
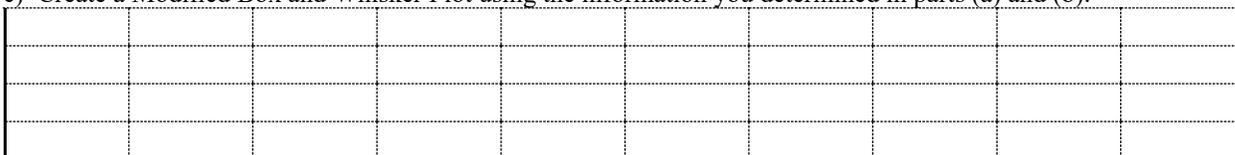
d) If the outlier were removed from the data set, how would this affect the mean of the data set? Explain your answer.

e) If the outlier were removed from the data set, how would this affect the standard deviation of the data set? Explain your answer.

31.   3 4 7 8 11   15   32

a) Use the data above to create a 5-number summary

Min	Q1	Med	Q3	Max

b) Determine the Interquartile Range (IQR), Lower Boundary, and Upper Boundary to determine and identify any outliers.

IQR_____

Lower Boundary_____

Upper Boundary_____

Outlier(s)_____

c) Create a Modified Box and Whisker Plot using the information you determined in parts (a) and (b).
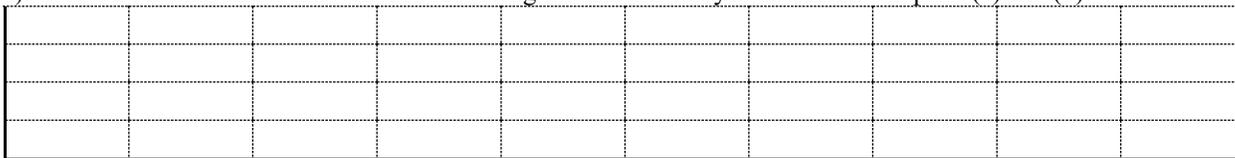
d) If the outlier were removed from the data set, how would this affect the mean of the data set? Explain your answer.

e) If the outlier were removed from the data set, how would this affect the standard deviation of the data set? Explain your answer.

32.   19   27   7   15   29   22   22   15   12   31

a) Use the data above to create a 5-number summary

Min	Q1	Med	Q3	Max

b) Determine the Interquartile Range (IQR), Lower Boundary, and Upper Boundary to determine and identify any outliers.

IQR_____

Lower Boundary_____

Upper Boundary_____

Outlier(s)_____

c) Create a Modified Box and Whisker Plot using the information you determined in parts (a) and (b).
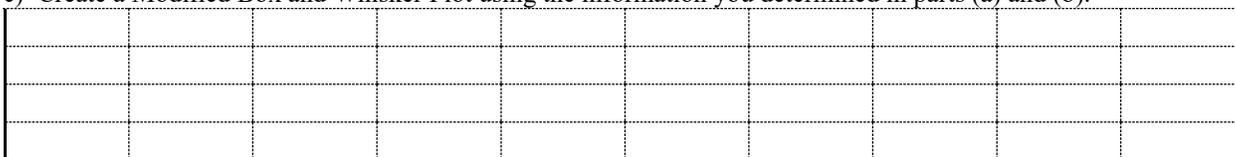
33.   150 customers bought lunch at the Food Court. The amounts spent were distributed normally with a mean of $5.50 and a standard deviation of $0.50. Label the mean and three standard deviations from the mean.
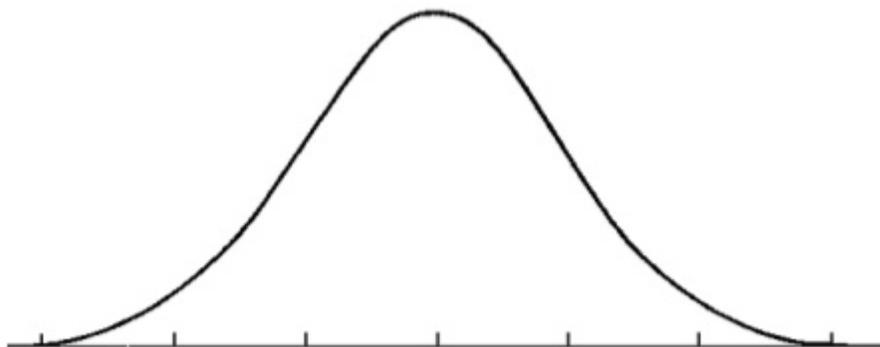
a) What percentage of customers spent between $4.50 and $6.00?

b) What percentage of customers spent between $5.50 and $6.00?

c) What percentage of customers spent between $4.50 and $6.50?

d) What percentage of customers spent above $5.00?

e) What percentage of customers spent above $7.00?

f) Approximately how many customers spent between $4.50 and $6.50?

g) Approximately how many customers spent between $5.00 and $6.00?

h) Approximately how many customers spent between $4.50 and $5.50?

i) Approximately how many customers spent between $6.50 and $7.00?

j) Approximately how many customers spent below $6.00?

k) (Without Calculator) If Chi Spurger spends $5.15, his z-score would be between the integers ____ and ____ .

l) (With Calculator) Now calculate the exact z-score of Chi Spurger's food purchase of $5.15 at the Food Court.

m) (Without Calculator) Case Udia has a z-score of 2.4, then he spent between ____ and ____ .

n) (With Calculator) Now calculate exactly how much money Case Udia spent with a z-score of 2.4 .

Math 3 Unit 5   Review Worksheet

34. A standardized test has a mean score of 525 with a standard deviation of 50. A student scored 320 on this test. Determine the student's z-score.

35. The weight of chocolate bars from a particular chocolate factory has a mean of 8 ounces with standard deviation of 0.1 ounce. What is the z-score corresponding to a weight of 8.17 ounces?

36. A sample of size 300 from a normal population has a mean of 265 and a standard deviation of 40. Using the Empirical Rule, about how many items of the sample will be above 305? How many will be below 225?

37. A married couple are employed by the same company. The husband works in a department for which the mean hourly rate is $15 and the standard deviation is $2. His wife is employed in a department where the mean rate is $18 with a standard deviation of $3. Relative to their departments, which is better paid if the husband earns $18 and the wife earns $19.50? Assume the distribution for both rates is normal.

38. Books in the county library are found to have average length of 350 pages with standard deviation of 100 pages. If the z-score for a copy Kafka's Metamorphosis is -2.7, what is its corresponding number of pages?

39. The average Rainbow Trout in Shaver Lake weighs 3.5 kg with a standard deviation of 1.1 kg. In Pine Flat Reservoir, the average weight of the Large Mouth Bass is 5.4 kg with a standard deviation of 0.9 kg. Michelle catches a 2.2 kg Rainbow at Shaver and Eddie catches a 4.1 kg Bass at Pine Flat. Relative to the other fish in the lakes, who had the better catch? Assume the distribution of the fish is normal.

40. Suppose Max Stout owes money on 7 different credit cards with debts between $550 and $7,800. He then decides to use a brand new credit card to purchase a self-help program entitled "Don't Buy Stuff you Cannot Afford" for $9,000. Select whether the value of each statistic, for the debt he owes on each credit card, increases, decreases, or cannot be determined.

	Increases	Decreases	Cannot Be Determined	Stays the Same
Mean				
Median				
Mode				
Standard Deviation				

41. Given the following 5-Number Summary for a set of test scores. Let the class size for the data equal 29.

Min	Q1	Med	Q3	Max
41	54	60	71	95

   a) Do any outliers exist in the class? Provide mathematical proof of your answer.

   b) Does the value 60 need to be an actual data value in the original set of test scores? Explain your answer.

42. Given the following 5-Number Summary for a set of test scores. Let the class size for the data equal 34.

Min	Q1	Med	Q3	Max
60	73	78	81	92

   a) Do any outliers exist in the class? Provide mathematical proof of your answer.

   b) Does the value 78 need to be an actual data value in the original set of test scores? Explain your answer.

43. The Juarez family took a 10 day vacation visiting various locations in California. Their hotel bills ($) were as follows:

   88   95   115   140   140   172   176   189   194   370

a) Find the 5-number summary

Min	Q1	Median	Q3	Max

b) Check for outliers and make a box and whisker plot (identify outliers, if any)

   Show work here:

50   75   100   125   150   175   200   225   250   275   300   325   350   375

c) Find the mean and the standard deviation of the 5 lower hotel bills: 88 95 115 140 140

   $\underline{x}$ μ $\underline{(x - \mu)}$ $\underline{(x - \mu)^2}$

                                                              Mean =

                                                              Std. Dev. =

d) In part b), you hopefully discovered an outlier. Remove the outlier and find the mean and median.

44. While in Santa Cruz, the Juarez family stayed at a hotel which was across the street from a large Chevrolet dealership. Of all the new vehicles on the lot, the mean and standard deviation selling price **of the cars** was $27,500 and $1,500. The mean and standard deviation **of the trucks** was $31,000 and $2,000. Both distributions follow a <u>Normal model</u>.

a) What is the standard score (z-score) of a truck which sells for $34,800?

b) If a car has a z-score of -1.2, what is it's selling price?

c) Draw and label a normal distribution graph for the trucks and then use it to answer part h)

d) Regarding all of the <u>trucks</u> at the dealership, what % are priced…….

More than $33,000_____

Between $27,000 and $33,000_____

Less than $27,000_____

The middle 95% of all trucks are priced between

_____   and   _____

e) If there are 51 trucks which are priced between $29,000 and $31,000, how many trucks are for sale at the dealership?

f) Relatively speaking, what would be more expensive a $36,000 car or a $36,000 truck? Why?

45. Carla, the neighbor who fed the Juarez's fish while they were on vacation, is a Human Relations director of a large multiple location department store. She wishes to seek information from the employees regarding a new health care plan. After reading Carla's plan on how to gain information from the employees, identify the sample method as:

> A-Simple Random Sample
>
> B-Convenience
>
> C-Systematic
>
> D-Self Selected

a) Carla decides to create a website and allow any employee to log on and volunteer their opinion regarding the new health care plan _____

b) Carla knows that each employee has a unique employee ID #. She uses a Random Digit Table to select 50 employees who will receive a survey regarding the new health care plan _____

c) Carla has an alphabetical listing of all employees. She sends a survey to every 10th employee on the list _____

d) The corporate offices are located next to the largest of the store locations. Carla knows that the employees of that store report to work between 8:45 and 9:00 each morning. She waits at the employee entrance and hands out a survey to the first 30 people who report to work one morning _____

46. A large airline has several daily flights from Los Angeles to New York. The average number of passengers for each flight follows a Normal model with a mean of 188 and a standard deviation of 10.

a) Use the Empirical Rule to help you determine the percentage of flight which carry:

More than 208 passengers_____

Between 178 and 218 passengers_____

95% of all flights carry between_____and_____ passengers

b) The flight you are on has a total of 174 passengers. Calculate <u>and explain</u> the z-score of your flight.

c) Another flight had a z-score of 2.7. How many passengers were on that flight?

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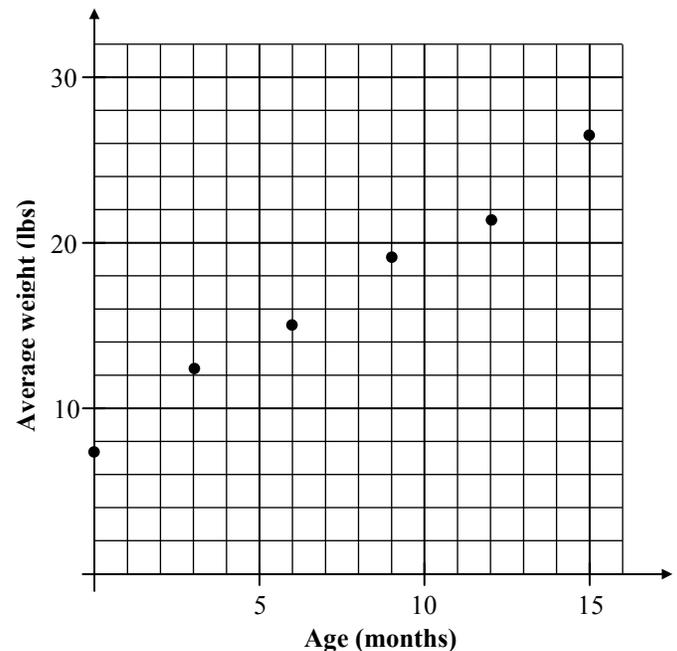
(c) Using the linear regression command on your calculator, find the equation of the best fit line

(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) $y = 6.5x + 3.9$

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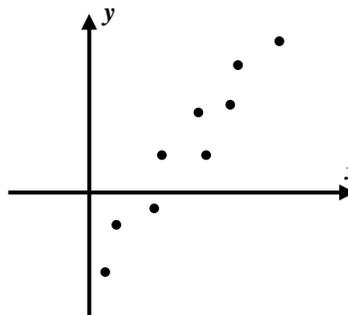
3. A line of best fit was created for a data set that only included values of x on the interval $12 \le x \le 52$. For which of the following values of x would using this model represent extrapolation?

   (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 I

In the last two lessons we fit **bivariate data sets** with **lines of best fit**. Sometimes, though, linear models are not the best choice. We can fit data with all sorts of curves, the most common of which are **linear**, **exponential**, and **quadratic**. But, there are many other types. Before we look at exponential and quadratic regression, recall the general shapes of these two types of functions.

EXPONENTIAL AND QUADRATIC GRAPHS

EXPONENTIAL GRAPHS   **QUADRATIC GRAPHS**

***Exercise* #1:** For each scatterplot shown below, determine if it is best fit with a linear, exponential, or quadratic function. Draw a curve of best fit depending on your choice.

(a)

Type: _____

(b)

Type: _____

(c)

Type: _____

(d)

Type: _____

(e)

Type: _____

(f)
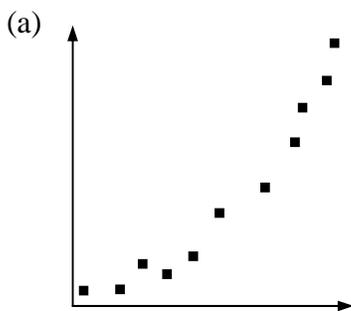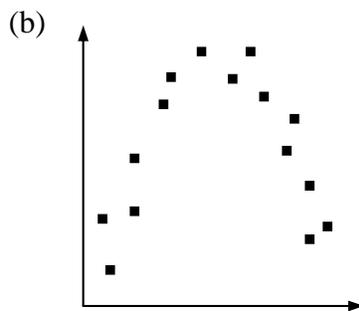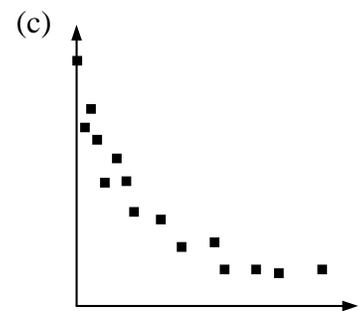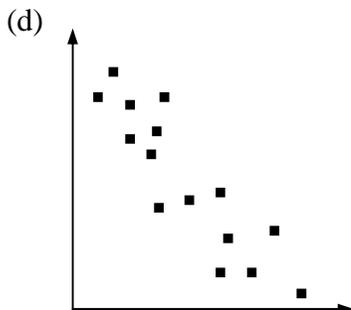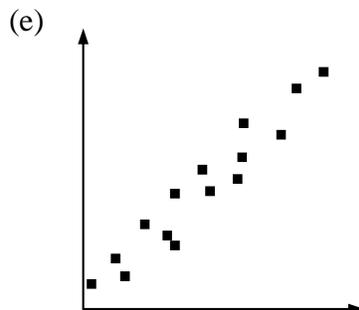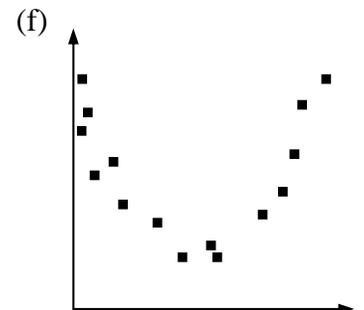
Type: _____

Our calculators can produce equations for **exponentials of best fit** and **quadratics of best fit** (along with a lot of other types of curves).

Exercise #2: Biologists are modeling the number of flu cases as it spreads around a particular city. The total number of cases, y, was recorded each day, x, after the total first reached 16. The data for the first week is shown in the table below.

x, days	0	1	3	4	6	7
y, cases	16	18	22	25	33	35

(a) Use your calculator to find the **exponential regression equation** for this data set in the form $y = a(b)^x$ Round all parameters to the nearest *hundredth*.

(b) Based on the regression equation, how many total cases of flu will there be after two weeks?

(c) According to your model, by what percent are the flu cases increasing on a daily basis?

(d) Hospital officials will declare an emergency when the total number of cases exceeds 200. On what day will they need to declare this emergency?

So, really, regression, as mysterious as it may be, is all about finding the best version of whatever curve we think fits the data best.

Exercise #3: The cost per widget produced by a factory generally drops as more are produced but then starts to rise again due to overtime costs and wear on the equipment. Quality control engineers recorded data on the cost per widget compared to the number of widgets produced. Their data is shown below.

Number of widgets, x	35	88	110	135	154	190
Cost per widget, y	9.32	2.63	1.42	1.32	2.12	5.50

(a) Why should a quadratic model be considered for this data set as opposed to linear or exponential?

(b) Use your calculator to create a scatterplot of this data to verify its quadratic nature.

Name: _____   Date: _____

OTHER TYPES OF REGRESSION
COMMON CORE ALGEBRA I HOMEWORK

FLUENCY

1. For each scatterplot below, determine the best type of regression from: linear, exponential, or quadratic. Draw a representative curve (line, exponential, or parabola) through the data.
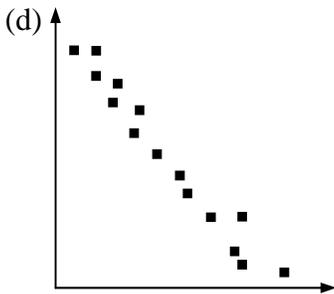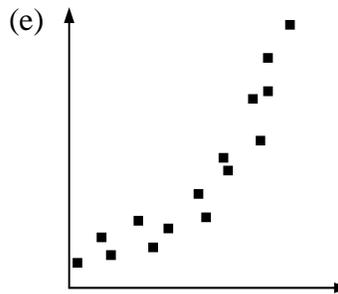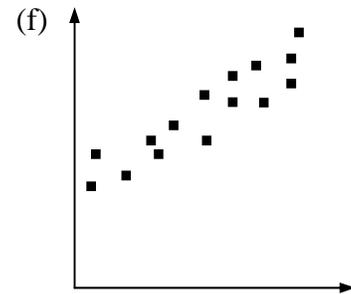
(a)

Type: _____

(b)

Type: _____

(c)

Type: _____

(d)

Type: _____

(e)

Type: _____

(f)
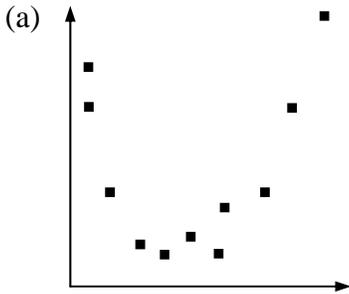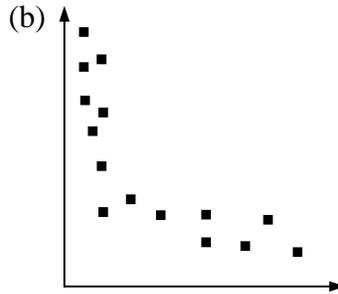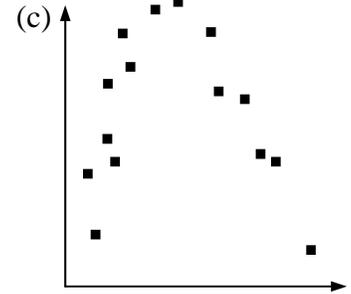
Type: _____

2. Given the scatterplot below, which of the following equations would best model the data? Explain your choice.

(1) $y = -3x + 6$ (3) $y = -4x^2 + 20x + 3$

(2) $y = 6(2)^x$ (4) $y = 2x^2 - 6x + 4$
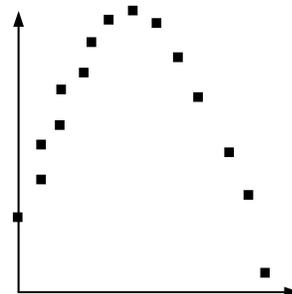
APPLICATIONS

3. A marketing company is keeping track of the number of hits that a website receives on a daily basis. Their data for the first two weeks is shown below. A scatterplot of the data is also shown.

Days	Hits
0	120
3	145
5	162
10	220
14	270

Daily Hit Count for Site

Days After the Website Launched
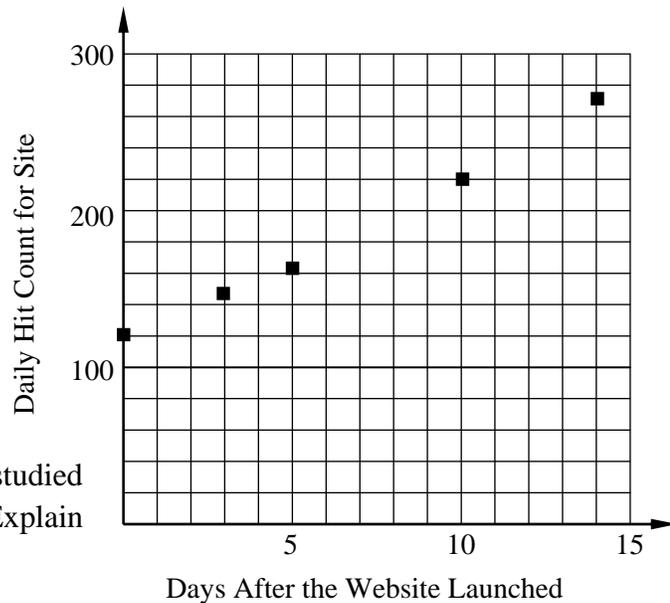
(a) Of the three types of regression we have studied which seems least likely to fit this data? Explain your choice.

(b) Find a linear equation, in the form $y = ax + b$, that best models this data and an exponential equation, in the form $y = a(b)^x$ that best models this data. Round all parameters to the nearest *hundredth*.

Linear Model   **Exponential Model**

(c) How close are the two model's outputs when $x = 10$? Show the values you find.

(d) How close are the two model's outputs when $x = 30$? Show the values that you find.

(e) Which model will predict faster growth of website hits over time? Explain your answer. You may want to experiment by graphing both models.

QUANTIFYING PREDICTABILITY
COMMON CORE ALGEBRA I

In the last few lessons we have worked with generating **lines and curves of best** fit for **bivariate** data sets. In every circumstance, though, the data did not fall on a straight line or on a perfect curve. We have never answered the question of how well specifically a **linear model** does in predicting the **correlation** between the two variables.

Exercise #1: Your teacher will explain how to ensure that your calculator has its "r-value" on. Since this varies by graphing calculator, write down the procedure below if necessary.

Exercise #2: In the following exercises four data sets with equal *x*-values are given to illustrate different types of **positive correlations**. For each, enter the data, observe the scatter plot, and record the *r*-value, known as the **correlation coefficient**, for a **linear fit** to the nearest *thousandth*.

(a)

x	2	5	8	11	15	18
y	4	13	22	29	43	52

(b)

x	2	5	8	11	15	18
y	16	14	22	41	37	51

(c)

x	2	5	8	11	15	18
y	18	8	41	28	62	44

(d)

x	2	5	8	11	15	18
y	44	51	30	55	45	47

(d) How does the **correlation coefficient** quantify the fit of a positive correlation?

Exercise #3: The following data set is that of two variables that have a **negative correlation**. Enter the data, produce the scatter plot, and record the *r*-value. How is the negative correlation reflected in the *r*-value?

x	2	5	8	11	15	18
y	52	47	28	32	25	10

Exercise #4: Given the scatter plot shown below, which of the *r*-values would most likely represent the correlation between the two variables? Explain your choice.
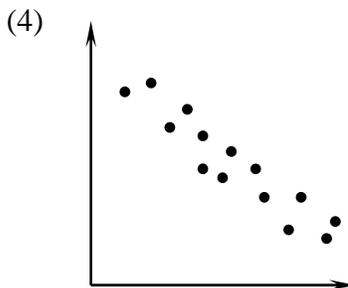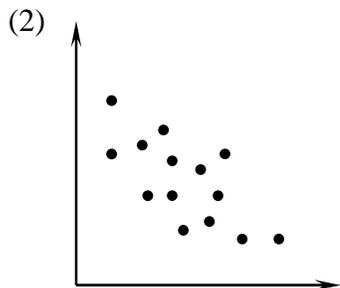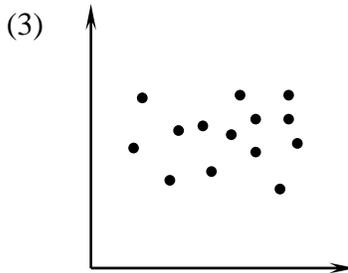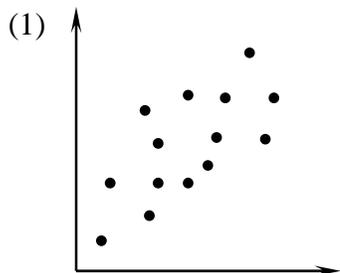
(1) $r = 0.88$ (3) $r = 1$

(2) $r = 0.28$ (4) $r = -0.94$
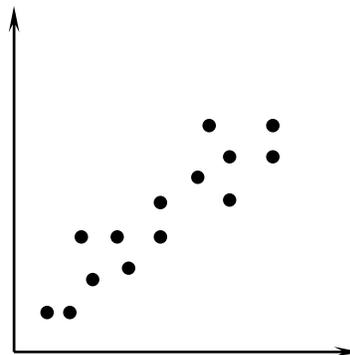
Exercise #5: Which of the following scatter plots would have a correlation coefficient closest to -1?

(1)

(3)

(2)

(4)

Exercise #6: There are two primary types of crude oil sold in the world, West Texas Intermediate (WTI) and Brent Crude. Each is priced differently on a daily basis and each has a correlation with the average price per gallon for unleaded gasoline. The two linear regression models, along with their *r*-values, are shown below. Give a prediction for the price per gallon of unleaded gasoline, *y*, on a day when the price for WTI is $103 and the price for Brent is $109, *x*. Which model did you choose and why?

Brent Crude: $y = 0.028x + 0.71$, $r = 0.973$ WTI Crude: $y = 0.031x + 0.67$, $r = 0.924$

Name: _____   Date: _____

QUANTIFYING PREDICTABILITY
COMMON CORE ALGEBRA I HOMEWORK

1. Below there are six scatter plots, six correlation coefficients, and six terms. Match the appropriate *r*-value with the scatter plot it most likely corresponds to. Then match the term you think is most appropriate to the *r*-value as well (not to the graph).
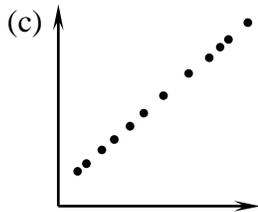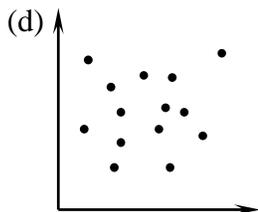
(a)

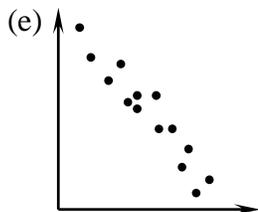$r = 1.0$

Weak Negative

(b)
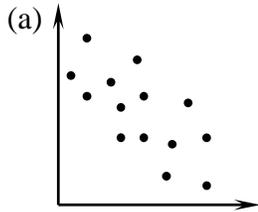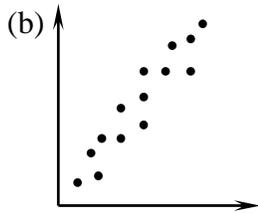
$r = 0.35$

Perfect Positive

(c)

$r = -0.82$
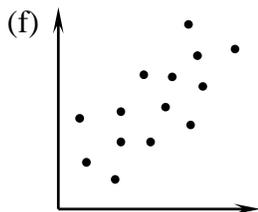
Strong Positive

(d)

$r = 0$

Weak Positive

(e)

$r = -0.56$

Moderate Negative

(f)

$r = 0.93$

No Correlation

2. A solar power company is trying to correlate the total possible hours of daylight (simply the time from sunrise to sunset) on a given day to the production from solar panels on a residential unit. They created a scatter plot for one such unit over the span of five months. The scatter plot is shown below.

The equation line of best fit for this bivariate data set was:

$y = 2.26x + 20.01$.

(a) How many kilowatt hours would the model predict on a day that has 14 hours of possible daylight?
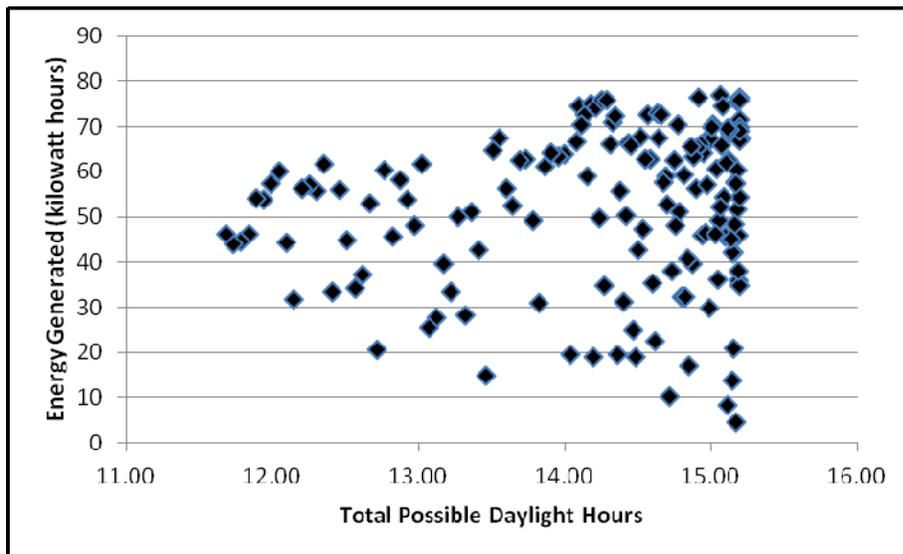
(b) To the nearest tenth of an hour, how many hours of possible daylight would be needed to produce 50 kilowatt hours of energy?

(c) The correlation coefficient for this regression was $r = 0.134$. Would you characterize this as strongly positive, moderately positive, or a weakly positive correlation? Explain.

(d) Based on (c), do you have confidence in the model to accurately predict the energy production based on the total possible daylight hours? Explain.

(e) What environmental factors might contribute to the "noise" in the data? Noise are factors that prevent the correlation from being perfect.