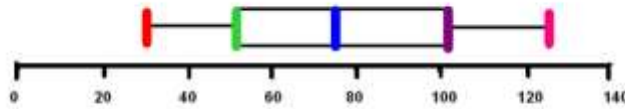# Statistics Mini Packet

Box and Whisker Plots are graphs created to display data (very similar to a histogram) at a quick glance. Box and Whisker Plots can also be called Box Plots. There are 5 key pieces.

Annual snow depth at Mathsville Ski Resort

Lowest Value:
The smallest number that is in the set of data.

Highest Value:
The largest number that is in the set of data.

Lower (First) Quartile:
The median of the lower half of the given

Median (Second Quartile):
The middle of the given data Represents 50%

Upper (Third) Quartile:
The median of the upper half of the given data. Represents 75%

Range:
The amount of numbers represented in the set of data. To identify the range, you will subtract the lowest value from the highest value.
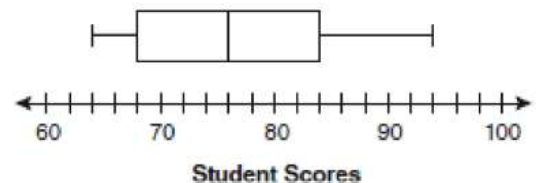

Exercise #1: Given the following box and whisker plot what is:

a) the lowest value:____**55**____

b) the highest value:____**100**____

c) the lower quartile:____**60**____

d) the upper quartile:____**85**____

e) the median value:____**75**____
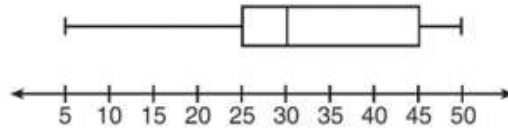
f) What is the range? ___**100 - 55 = 45**____
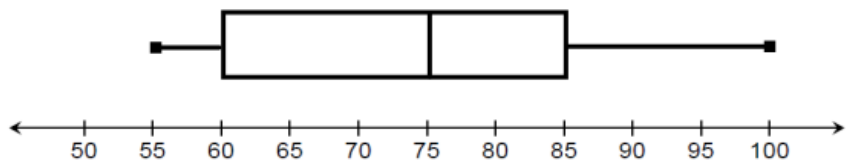
Exercise #2: The box-and-whisker plot below represents students scores on a recent English test. What is the value of the lower quartile?
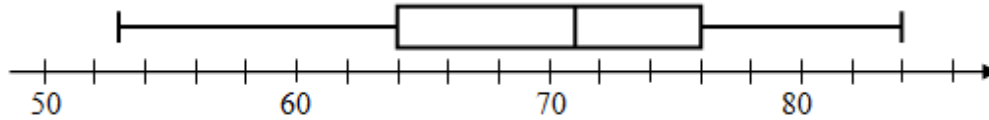
(1) 68 (2) 76

(3) 84 (4) 94

Exercise #3: In the box plot provided, what is the value for the 2nd quartile?

(1) 25   (2) 30

(3) 45   (4) 50

Exercise #4: Twenty of Mr. Smith's math students recently took a quiz. The results of this quiz are shown in the following box-and-whiskers diagram. Assume that all scores are whole numbers.

(a) What was the median score on Mr. Smith's math quiz?

(b) What was the range of the scores on Mr. Smith's math quiz?

(c) What score was greater than or equal to 75% of all the other scores on this quiz?

(d) Mr. Smith regularly sets the passing grade on his quizzes to be the score of the lower quartile. What is the passing grade on this quiz?
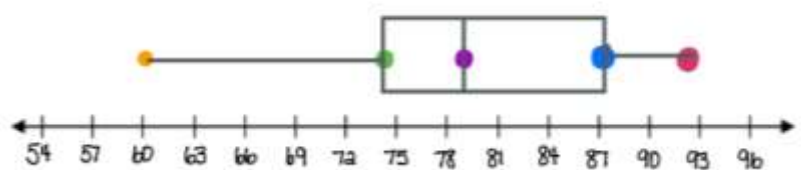
Exercise #5: Shown below are the scores 12 students received on a math quiz. Create a box and whisker plot to represent the data.   64, 80, 88, 78, 60, 92, 84, 76, 86, 78, 72, and 90

How to use the calculator to find **Measures of Central Tendencies**:

1.) Press Stat

2.) Press Edit (if you have data in the list already, use the up arrow to highlight L1, then press clear and enter).

3.) Type in your data set. Be very careful and take your time!

4.) Press Stat

5.) Scroll to the right to (Calc)

6.) Choose option #1 (1-Var Stats)

7.) Press Enter three times

8.) Use the screen information to answer each question.

NORMAL FLOAT AUTO a+bi RADIAN MP

1-Var Stats

$\bar{x}=79$ MEAN
$\Sigma x=948$
$\Sigma x^2=75984$
Sx=9.963570006 SAMPLE SIZE STANDARD DEVIATION
σx=9.539392014 POPULATION STANDARD DEVIATION
n=12 SAMPLE SIZE (THE NUMBER OF ITEMS YOU INPUT)
↓Q_1=74 FIRST QUARTILE
■Med=79 MEDIAN VALUE
Q_3=87 THIRD QUARTILE
maxX=92 LARGEST VALUE

Min = 60
Q1 = 74
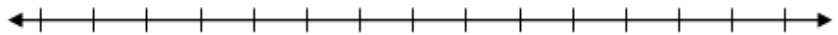Median = 79
Q3 = 87
Max = 92

2

Exercise #6: Shown below are the scores 16 students received on a math quiz.
52, 60, 66, 66, 68, 72, 72, 73, 74, 75, 80, 82, 84, 91, 92, 98

(a) What is the median of this data set?

(b) Find the range of the data set.

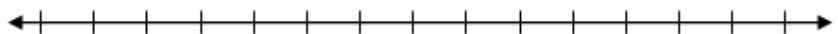(c) Construct a box plot on the number line given below.

Exercise #7: The ages of the 15 employees of the Red Hook Curry House are given below.
16, 17, 17, 18, 19, 22, 25, 26, 29, 33, 33, 37, 40, 42, 44

(a) What is the first quartile of this data set?

(b) Find the maximum value of the data set.

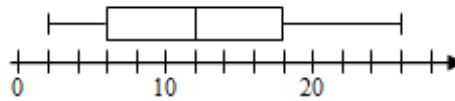(c) Construct a box plot on the number line given below.

____ 1. Which of the following data sets, given in ascending order, has the greatest range?
    (1) $\{3, 4, 7, 10, 18\}$ (3) $\{-2, 5, 8, 11, 26\}$

    (2) $\{65, 66, 70, 72\}$ (4) $\{-5, -2, 4, 7, 10\}$
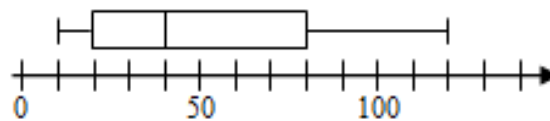

____ 2. Given the box plot shown below, which of the following represents the third quartile value for this data set?

    (1) 12   (3) 6

    (2) 18   (4) 19

____ 3. Given the box plot shown below, which of the following represents the range of this data set?

    (1) 110   (3) 60

    (2) 40   (4) 75

____ 4. According to the following box-and-whiskers diagram, which of the following values represents the lower
    quartile of this data set?

    (1) 20   (3) 28

    (2) 13   (4) 16
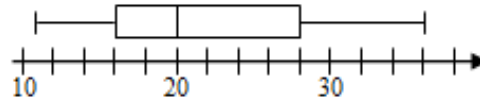
____ 5. Which of the following box-and-whiskers diagram represents a data set whose median value is equal to 65?

    (1)   (3)

    (2)   (4)
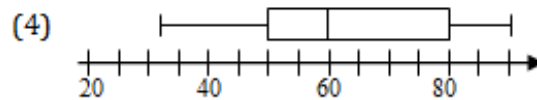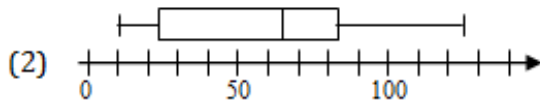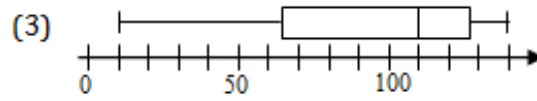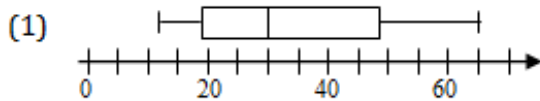
____ 6. The box-and-whisker plot below represents the math scores of 20 students. What percentage
    of the test scores are less than 72?
        (1) 25   (2) 50
        (3) 75   (4) 100

4
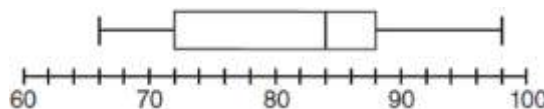
7. Mr. Ramirez gives a math test and records the grades of his 17 students as follows:

67, 72, 74, 74, 78, 80, 80, 82, 85, 85, 86, 87, 90, 92, 92, 95, 98

Create a box-and-whisker diagram of this data set below.

```
 ├───────┼───────┼───────┼───────┼───────┼───────┼───────┼───────┼──►
        60      70      80      90     100
```
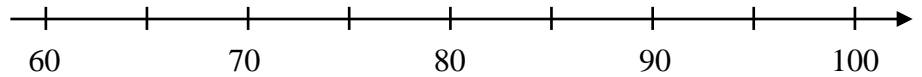
8. The speeds, in miles per hour, of 24 cars on a particular road are recorded and represented on the box-and-whiskers diagram shown below. Answer each of the following questions based on this diagram.

(a) What is the range of this data set?


```
 ├─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─►
      20       30       40       50
```
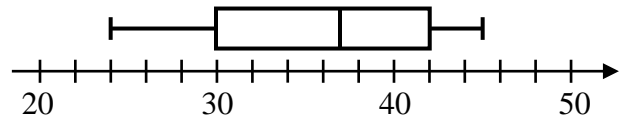
(b) What is the maximum speed of the 24 drives?   (c) How many drivers drove between 30 and 42 miles per hour?

(d) If the speed limit on this part of the road is 35 miles per hour, are more people speeding or are more people going below the speed limit? Justify your answer.

Review Section:

____ 9.) The equation $y = ax^2 + bx + c$ is graphed on the set of axes below. Based on the graph, what are the roots of the equation $ax^2 + bx + c = 0$?

[1] 0 and 5   [3] 1 and 5

[2] 1 and 0   [4] 3 and -4
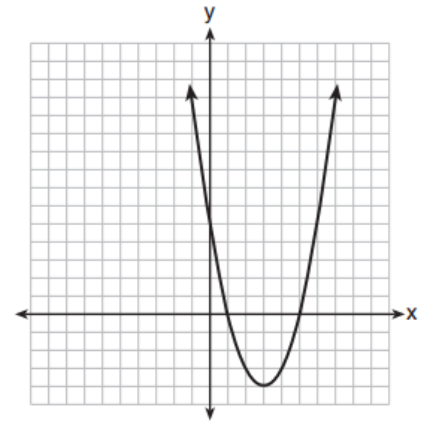
____ 10.) What is the solution of $4x - 30 \geq -3x + 12$?

[1] $x \geq 6$ [3] $x \geq -6$

[2] $x \leq 6$ [4] $x \leq -6$

11.) Solve the following system of equations algebraically for all values of x and y. ** **Hint: Quadratic Equations** **

$$y = x^2 + 2x - 8$$
$$y = 2x + 1$$

1.) 3

2.) 2

3.) 1

4.) 4

5.) 2

6.) 1

7.) Box Plot

8.) a.) range = 21 mph
    b.) 45 mph
    c.) 50% of the drives fall into this range, therefore 12 drivers
    d.) The median is 37 mph meaning that half of the drivers are going more than 37mph. Therefore, more people are speeding because the median is greater than the speed limit.

9.) 3

10.) 1

11.) $(-3, -5)$ and $(3,7)$

Name: _____   Date: _____ Period: _____
Algebra   Measures of Central Tendency   14C

In our day to day activities, we deal with many problems that involve related items of numerical information called *data*. *Statistics* is the study of sets of such numerical data. When we gather numerical data, besides displaying it, we often want to know a single number that is representative of the data as a whole. We call these types of numbers **measures of central tendency**. The two most common measures of central tendency are the **mean** and the **median**.

>   *Mean*: The mean is the average of the numbers: a calculated "central" value of a set of numbers.

>   *Median*: The Median is the *"middle"* of a sorted list of numbers.

>   *Mode*: The number which appears most often in a set of numbers.

*Example***:** Sara's test scores in mathematics were 64, 80, 88, 78, 60, 92, 84, 76, 86, 78, 72, and 90. Determine the mean, the median, and the mode of Sara's test scores.

How to use the calculator to find **Measures of Central Tendencies**:

1.) Press Stat
2.) Press Edit (if you have data in the list already, use the up arrow to highlight L1, then press clear and enter).
3.) Type in your data set. Be very careful and take your time!
4.) Press Stat
5.) Scroll to the right to (Calc)
6.) Choose option #1 (1-Var Stats)
7.) Press Enter three times
8.) Use the screen information to answer each question.

NORMAL FLOAT AUTO a+bi RADIAN MP

1-Var Stats
x̄=79   MEAN
Σx=948
Σx²=75984
Sx=9.963570006   SAMPLE SIZE STANDARD DEVIATION
σx=9.539392014   POPULATION STANDARD DEVIATION
n=12   SAMPLE SIZE (THE NUMBER OF ITEMS YOU INPUT)
minX=60   SMALLEST VALUE
↓Q₁=74   FIRST QUARTILE
■Med=79   MEDIAN VALUE
Q₃=87   THIRD QUARTILE
maxX=92   LARGEST VALUE

| The mean = 79 | The median = 79 | The mode = 78 |

Exercise #**1:** A survey was taken amongst 12 people on the number of passwords they currently have to remember. The results in ascending order are shown below. State the median number of passwords and the mean number of passwords (to the nearest tenth).

$$0, 1, 1, 1, 2, 2, 3, 3, 3, 3, 4, 6$$

Exercise #2: Students in Mr. Tobin's algebra class were trying to determine if people speed along a certain section of roadway. They collected speeds of 10 vehicles, as displayed in the table below.
(a) Find the mean and median for this data set.

Speed (mph)	Number of Cars
29	1
33	4
34	2
35	3

(b) The speed limit along this part of the highway is 34 mph. Based on your results from part (a), is it fair to make the conclusion that the average driver goes the speed limit on this roadway?

When conducting a statistical study, it is not always possible to obtain information about every person or situation to which the study applies. Unlike a census, in which every person is counted, some studies use only a **sample** or **portion** of the items being investigated. Whenever a sample is taken, it is vital that it be **fair**; in other words, the sample reflects the overall population.

Exercise #3: To determine which television programs are the most popular in a large city, a poll is conducted by selecting a sample of people at random and interviewing them. Outside which of the following locations would the interviewer be most likely to find a fair sample? Explain your choice and why the others are inappropriate.

(1) A baseball stadium

(3) A grocery store

(2) A concert hall

(4) A comedy club

9

Exercise #4: Truong is trying to determine the average height of high school male students. Because he is on the basketball team, he uses the heights of the 14 players on the team, which are given below in inches.

$$69, 70, 72, 72, 74, 74, 74, 75, 76, 76, 76, 77, 77, 82$$

(a) Calculate the mean and median for this data set. Round any non-integer answers to the nearest tenth.

(b) Is the data set above a **fair sample** to use to determine the average height of high school male students? Explain your answer.

Data sets can have members that are far away from all of the rest of the data set. These elements are called **outliers,** which can result in a mean that does not represent the true "average" of a data set.

Exercise #5: In Mr. Petrovic's Advanced Calculus, eight students recently took a test. Their grades were as follows:
$$45, 78, 82, 85, 87, 89, 93, 95$$

(a) Calculate the mean and median of this data set.   (b) What score is an **outlier** in this data set?

(c) Which value, the mean or the median, is a better measure of how well the average student did on Mr. Petrovic's quiz?

Name: _____   Date: _____ Period: _____
Algebra   Measures of Central Tendency   14C HW

____ 1.) The Student Government at Arlington High School decided to conduct a survey to determine where to go
          on a senior field trip. They asked students the following question: "Would you rather go to a sports event
          or to an IMAX movie?" At which of the following locations would they most likely get a fair sample?
          [1] The gym, after a game   [3] A randomly chosen study hall

          [2] The auditorium after a play   [4] At the Nature Club meeting.

____ 2.) What was the median high temperature in Middletown
          during the 7-day period shown in the table?

Daily High Temperature in Middletown	
Day	Temperature (°F)
Sunday	68
Monday	73
Tuesday	73
Wednesday	75
Thursday	69
Friday	67
Saturday	63

          [1] 69
          [2] 70
          [3] 73
          [4] 75

____ 3.) On an English examination, two students received scores of 90, five students received 85, seven students
          received 75 and one student received 55. The average score on this examination was:
          [1] 75   [3] 77

          [2] 76   [4] 79

____ 4.) What is the mean of the data in the accompanying table?
          [1] 11   [3] 15

          [2] 14.5   [4] 16

Scores (x_i)	Frequency (f_i)
25	3
20	2
11	5
10	4

5.) Given the following list of students' scores on a quiz:
              5, 12, 7, 15, 20, 14, 7
     Determine the median and the mode of these scores. The teacher decides to adjust these scores by adding three
points to each score. Explain the effect, if any, that this will have on the median and the mode of these scores.

____ 6.) Which of the following is true about the data set {3, 5, 5, 7, 9}?
      [1] median > range   [3] mean > median
      [2] median = mean   [4] median > mean

____ 7.) Which of the following data sets has a median of 7.5?
      [1] {6, 7, 8, 9, 10}   [3] {1, 3, 7, 10, 14}
      [2] {3, 5, 7, 8, 10, 14}   [4] {2, 7, 9, 11, 14, 17}

____ 8.) Mrs. Porter recorded her students' grades in the table provided. Which statement is true for the data?
      [1] mean > median > mode

      [2] mean > mode > median

      [3] mode > median > mean

      [4] median > mean > mode

Score	Frequency
96	2
92	5
88	3
84	2
78	4
60	1

9.) A survey is taken by an insurance company to determine how many car accidents the average New York City resident has gotten into in the past 10 years. The company surveyed 20 people who are getting off a train at a subway station. The following table gives the results of the survey.
(a) Calculate the mean, median and mode of the number of accidents of this data set.

Number of Accidents	Number of People
0	6
1	8
2	4
3	1
11	1

(b) Are there any outliers in this data set? If so, what data value?

(c) Which number, the mean or the median, better represents the number of accidents an average person in this survey had over this 10 year period? Explain your answer.

(d) Does this sample fairly represent the average number of accidents a typical New York City resident would get into over a 10 year period? Why or why not?

(e) Construct a dot plot that represents this data on the set of axes provided. Is this a symmetric plot? Explain your thinking.

Number of Traffic Accidents
0   2   4   6   8   10   12
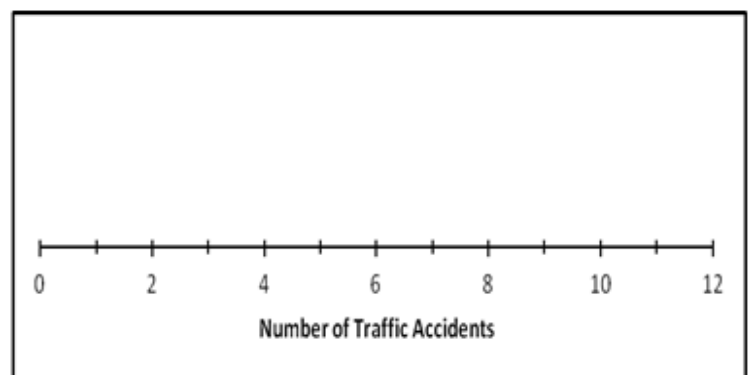
Review Section:

_____ 10.) What is an equation of the line that passes through the points (2,1) and (6,-5)?

[1] $y = -\frac{3}{2}x - 2$

[2] $y = -\frac{3}{2}x + 4$

[3] $y = -\frac{2}{3}x - 1$

[4] $y = -\frac{2}{3}x + \frac{7}{3}$

11.) Carla bought a dress at a sale for 20% off the original price. The sale price of the dress was $28.80. Find the original price of the dress.

12.) A DVD costs twice as much as a music CD. Jack buys 2 DVDs and 2 CDs and spends $45. Determine how much one CD costs. [Only an algebraic solution can receive full credit.]

Name: ___*Homework Answers*_____   Date: _____ Period: _____
Algebra   Measures of Central Tendency   14C HW

1.) [3]

2.) [1]

3.) [4]

4.) [3]

5.) 12 and 7. Both the median and the mode will increase.

6.) [3]

7.) [2]

8.) [3]

9.) a.) Mean = 1.5   Median = 1   Mode = 1
     b.) Yes, the one person who had 11 accidents
     c.) The median, 14/20 people had one accident or less, so the mean of 1.5 is too high
     d.) No, the sample is surveying people getting off a train at a subway station. Therefore, they are in a
         car less than someone asked in a different setting.
     e.) It is NOT symmetric.

10.) [2]

11.) The original price of the dress is $36.00

12.) One CD costs $7.50

Name: _____ Date: _____ Period: _____
Algebra   Variation within a Data Set   14D

Measures of central tendency give us numbers that describe the typical data value in a given data set. But, they do not let us know how much **variation** there is in the data set. Two data sets can have the same mean but look radically different depending on how varied the numbers are in the set.

Example: Sara's test scores in mathematics were 64, 80, 88, 78, 60, 92, 84, 76, 86, 78, 72, and 90. Determine the mean, the median, and the mode of Sara's test scores.

How to use the calculator to find **Measures of Central Tendencies**:

1.) Press Stat

2.) Press Edit (if you have data in the list already, use the up arrow to highlight L1, then press clear and enter).

3.) Type in your data set. Be very careful and take your time!

4.) Press Stat

5.) Scroll to the right to (Calc)

6.) Choose option #1 (1-Var Stats)

7.) Press Enter three times

8.) Use the screen information to answer each question.

NORMAL FLOAT AUTO a•bi RADIAN MP

1-Var Stats
$\bar{x}$=79   MEAN
Σx=948
Σx^2=75984
Sx=9.963570006   SAMPLE SIZE STANDARD DEVIATION
σx=9.539392014   POPULATION STANDARD DEVIATION
n=12   SAMPLE SIZE (THE NUMBER OF ITEMS YOU INPUT)
$\downarrow Q_1$=74   FIRST QUARTILE
$\blacksquare$Med=79   MEDIAN VALUE
Q_3=87   THIRD QUARTILE
maxX=92   LARGEST VALUE
minX=60   SMALLEST VALUE

> **1st Quartile = 74**   **3rd Quartile = 87**   **InterQuartile Range= Q₃ – Q₁ = 13**

Exercise #1: The scores of a recent math quiz are represented below.
(a) What are the values of the first quartile and the third quartile?
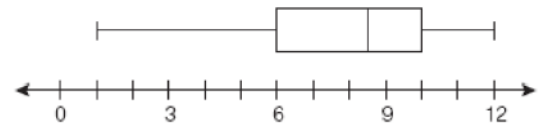
(b) What is the value of the interquartile range?

Exercise #2: The two data sets below each have equal means but differ in the variation within the data set. Use your calculator to determine (a) the mean of each set and (b) the **Interquartile Range (IQR)** of each data set.

Data Set #1: 3, 3, 4, 4, 5, 5, 6, 6, 7, 8, 8, 9, 9, 10, 10, 11, 11   **Data Set #2:** 5, 5, 6, 6, 7, 7, 8, 8, 9, 9

The interquartile range gives a good measure of how spread out the data set is. But, the best measure of variation within a data set is the **standard deviation**. The actual calculation of standard deviation is complex and we will not go into it here. We will rely on our _calculators_ for its calculation.

How to use the calculator to find **Measures of Central Tendencies**:

1.) Press Stat

2.) Press Edit (if you have data in the list already, use the up arrow to highlight L1, then press clear and enter).

3.) Type in your data set. Be very careful and take your time!

4.) Press Stat

5.) Scroll to the right to (Calc)

6.) Choose option #1 (1-Var Stats)

7.) Press Enter three times

8.) Use the screen information to answer each question.

NORMAL FLOAT AUTO a+bi RADIAN MP

1-Var Stats

$\bar{x}=79$ MEAN

$\Sigma x=948$

$\Sigma x^2=75984$

$Sx=9.963570006$ SAMPLE SIZE STANDARD DEVIATION

$\sigma x=9.539392014$ POPULATION STANDARD DEVIATION

$n=12$ SAMPLE SIZE (THE NUMBER OF ITEMS YOU INPUT)

$minX=60$ SMALLEST VALUE

$Q_1=74$ FIRST QUARTILE

$Med=79$ MEDIAN VALUE

$Q_3=87$ THIRD QUARTILE

$maxX=92$ LARGEST VALUE

STANDARD DEVIATION

The standard deviation of a data set tells us, on average, how far a data point is away from the mean of the data set. The larger the standard deviation, the greater the variation within the data set.

Exercise #3: The term "Snowstorms of Note" applies to all snowfalls over 6 inches. The snowfall amounts for "Snowstorms of Note" in Utica New York over a four-year period are as follows:

7.1, 9.2, 8.0, 6.1, 14.4, 8.5, 6.1, 6.8, 7.7, 21.5, 6.7, 9.0, 8.4, 7.0, 11.5, 14.1, 9.5, 8.6

What are the mean and population standard deviation for these data, to the nearest hundredth?

[1] mean = 9.46; standard deviation = 3.74   [2] mean = 9.46; standard deviation = 3.85

[3] mean = 9.45; standard deviation = 3.74   [4] mean = 9.45; standard deviation = 3.85

Population vs. Sample

When we are working with every possible data point of interest, we call this a **population** and therefore use the **population standard deviation** σ. When we have only a small portion of all possible values, we call this a **sample** and therefore we use the **sample standard deviation, s.** The formulas for these two differ slightly, so their values are often very similar, but slightly different.

Exercise #4: What is the sample standard deviation of the data set in the table, rounded to the nearest tenth?

[1] 12.5

[2] 12.8

[3] 17.1

[4] 18.7

Scores	Frequency
50	1
60	2
70	7
80	6
90	3
100	2

Exercise #5: The accompany frequency table shows the scores on a classroom test. What is the population standard deviation for this set of scores?

x_i	f_i
100	7
90	10
80	4
70	4

    [1] 10.2

    [2] 10.4

    [3] 25

    [4] 88

Exercise #6: Which of the following data sets would have a standard deviation (population) closest to zero?

    [1] $\{-5, -2, -1, 0, 1, 2, 5\}$ [3] $\{11, 11, 12, 13, 13\}$

    [2] $\{5, 8, 10, 16, 20\}$ [4] $\{3, 7, 11, 11, 11, 18\}$

____ 1.) Rosario and Enrique are in the same mathematics class. On the first five tests, Rosario received scores
of 78, 77, 64, 86, and 70. Enrique received scores of 90, 61, 79, 73, and 87. How much higher was
Enrique's average than Rosario's average?
[1] 15 points
[2] 2 points
[3] 3 points
[4] 4 points

____ 2.) The table below shows the first-quarter averages for Mr. Harper's statistics class.
What is the population variance for this set of data?
[1] 8.2
[2] 8.3
[3] 67.3
[4] 69.3

Statistics Class Averages

Quarter Averages	Frequency
99	1
97	5
95	4
92	4
90	7
87	2
84	6
81	2
75	1
70	2
65	1

____ 3.) The table displays the number of siblings of each of the 20 students in a class.
What is the population standard deviation, to the nearest hundredth, for this group?
[1] 1.11
[2] 1.12
[3] 1.14
[4] 1.15

Number of Siblings	Frequency
0	2
1	5
2	7
3	4
4	2

____ 4.) For the data shown in the dot plot, which of the following is closest to its population standard deviation?
[1] 2.7
[2] 4.2
[3] 3.3
[4] 5.8
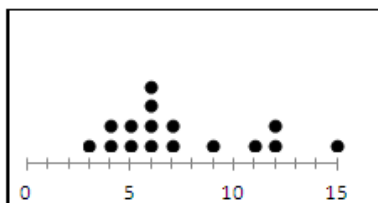
____ 5.) What is the interquartile range of the data represented in the box plot?
[1] 24
[2] 14
[3] 8
[4] 12
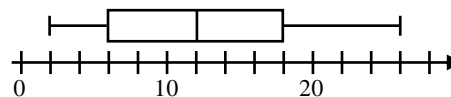
6.) The ages of ten teachers at George Washington Elementary School are: 33, 23, 36, 29, 36, 36, 33, 29, 36, 29.
Determine the population standard deviation of these ages to the nearest tenth.

7.) We are going to revisit our survey of households that have video enabled devices (televisions, smart phones, tablets, etcetera). Recall that two surveys were done, each with 30 participants. In the first case (Survey A), the survey was random, in the second case (Survey B), the survey only included families with at least one teenager. The dot plots of the results are shown below.

Number of Video Enabled Devices…

Survey A

Number of Video Enabled Devices…
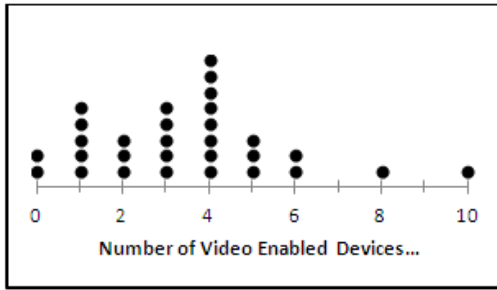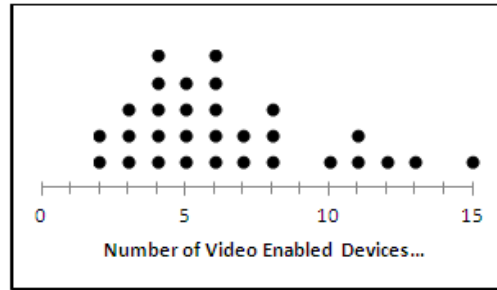
Survey B

(a) Calculate the mean number of devices, the interquartile range and the sample standard deviation of both sets of data. Round all non-integer answers to the nearest tenth.

Survey A Statistics: **Survey B Statistics:**

(b) Which of these two survey data sets had the greatest variation in the data? Explain based on the statistics you found in part (a).

(c) How many of the 30 values in Survey B fall within one standard deviation of the mean? To do this calculation, add the standard deviation and subtract the standard deviation from the mean and then count the number of values between the results of this addition and subtraction.

Review Section:

____ 7.) The expression $\dfrac{24x^6y^3}{-6x^3y}$ is equivalent to:

    [1] $-4x^2y^3$ [3] $-4x^9y^4$

    [2] $-4x^3y^3$ [4] $-4x^3y^2$
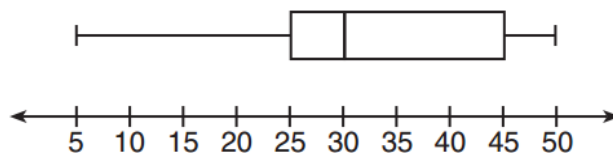
____ 8.) In the box and whisker plot, what is the 2nd quartile?

    [1] 25

    [2] 30

    [3] 45

    [4] 50

9.) The equation $6x + 5y = 300$ and $3x + 7y = 285$ represent the money collected from selling gift baskets in a school fundraising event. If x represents the cost of each snack gift basket and y represents the cost for each chocolate gift basket, what is the cost for each chocolate gift basket? **Don't forget your format – Let Solve Check**

20

Algebra *Homework Answers*

1.) [3]

2.) [1]

3.) [2]

4.) [3]

5.) [4]

6.) Population standard deviation = 4.2

7.) (a) A: mean = 3.4   IQR = 2   Sample SD = 2.3
          B: Mean = 6.4   IQR = 4   Sample SD = 3.3
   (b) Survey B; the interquartile range and the standard deviation are higher values, therefore they are farther
        from the mean and have the greatest variation.
   (c) 19 devices.

7.) [4]

8.) [2]

9.) A chocolate gift basket will cost $30.00

So far we have worked with **quantitative data** for a **single variable**, for example weight of baby chicks or number of video enabled devices. We can also work with **categorical data** or data that shows how many things **surveyed** fall into a **given category**.

Although surveys of data that contain only one category are interesting, statisticians are often interested in how responses to two categories relate to one another. For example, we may want to know how a person's gender (one category) affects what profession (a second category) they would prefer when they grow up. We may want to know if a person's hair color (one category) has any relationship to their eye color (a second category). This type of data is summarized in a **two-way frequency table**.

*Exercise #1***:** This table gives data on the number of tickets sold during three days at a movie theater.
(a) Find the total number of tickets sold on Saturday.

(b) Find the total number of tickets sold to Seniors.

	Friday	Saturday	Sunday
Adults	102	156	123
Seniors	34	98	102
Children	21	185	145

(c) Find the total number of tickets sold over the three days.

*Exercise #2***:** A Mathematics Class recorded their hair color and eye color which are shown in the two-way frequency table below.
(a) Find the total number of students with Blonde hair.

(b) Find the total number of students with Green eyes.

Hair Color

		Black	Blonde	Red
Eye Color	Blue	3	4	1
	Brown	5	2	0
	Green	1	1	3

(c) Find the total number of students in the Mathematics class.

We would like to understand **associations** or **trends** within the data set, i.e. would a response to one category tell us something about the response to the other category?

Exercise #3: This table gives data on the number of tickets sold during three days at a movie theater.
(a) What is the frequency of attendees at the show on **Saturday** to be **adults**?

$$\frac{Adults}{Total\ on\ Saturday} = \frac{156}{156+98+185} = \frac{156}{439}$$

	Friday	Saturday	Sunday
Adults	102	156	123
Seniors	34	98	102
Children	21	185	145

(b) What is the percentage of **seniors** who attended on **Friday**?

$$\frac{Friday}{Total\ of\ Seniors} = \frac{34}{34+98+102} = \frac{34}{234} = 0.1452991453 = 15\%$$

Exercise #4: A Mathematics Class recorded their hair color and eye color which are shown in the two-way frequency table below.
(a) What is the frequency of having Red Hair and Green Eyes.

Hair Color

Eye Color		Black	Blonde	Red
	Blue	3	4	1
	Brown	5	2	0
	Green	1	1	3

(b) What is the frequency of having Green Eyes and Red Hair.

(c) What percent of students have black hair and blue eyes?

(d) What percent of students have brown eyes and blonde hair?

Exercise #5: A survey of 52 graduating seniors was conducted to determine if there was a connection between the gender of the student and whether they were going on to college.

(a) What percent of females are going to college?

	Male	Female
Going to College	16	13
Not Going to College	14	9

(b) What is the frequency of not going to college as a male?

(c) What is the frequency of someone who is female and going to college?

(d) What is the percent of male students going to college?

Exercise #6: A radio state did a survey to determine what kind of music to play by taking a sample of middle school, high school, and college students. They were asked which of three different types of music they prefer on the radio: hip-hop, alternative, or classic rock. The results are summarized in the table.

(a) What percentage of college students prefer classic rock?

	Hip-Hop	Alternative	Classic Rock
Middle School	28	18	4
High School	22	22	6
College	16	20	14

(b) What is the frequency of middle school students who listen to hip-hop?

(c) What percentage of alternative music is listened to by High School students?

(d) What is the frequency of college students who listen to alternative music?

*Exercise #7***:**   A statistics class surveyed some students during one lunch period to obtain opinions about television programming preferences. The results of the survey are summarized in the table. Based on the sample, predict how many of the school's 351 males would prefer comedy. Justify your answer with work.

Programming Preferences

	Comedy	Drama
Male	70	35
Female	48	42

Name: _____   Date: _____ Period: _____
Algebra   Two Way Frequency Tables   14E HW

____ 1.) A public opinion poll was taken to explore the relationship between age and support for a candidate in an election. The results of the poll are summarized in the table. What percent of the 21 – 40 age group was for the candidate?

[1] 15
[2] 25
[3] 40
[4] 60

Age	For	Against	No Opinion
21–40	30	12	8
41–60	20	40	15
Over 60	25	35	15

2.) The school newspaper surveyed the student body for an article about club membership. The table below shows the number of students in each grade level who belong to one or more clubs. If there are 180 students in the 9th grade, what percentage of 9th grade students belong to more than one club?

	1 Club	2 Clubs	3 or More Clubs
9th	90	33	12
10th	125	12	15
11th	87	22	18
12th	75	27	23

3.) This table gives data on the number of tickets sold during three days at a movie theater.
(a) What is the frequency of attendees at the show on Sunday to be children?

	Friday	Saturday	Sunday
Adults	102	156	123
Seniors	34	98	102
Children	21	185	145

(b) What is the percentage of adults who attended on Saturday?

(c) What is the percentage of seniors who attended the show on Friday?

4.) A survey was done to determine the relationship between gender and subject preference. A total of 56 students were surveyed to determine if they liked math, English, social studies, or science as their favorite subject. The results were then broken down based on whether the respondent was male or female.

(a) What is the frequency of being a male who likes History?

	Math	English	History	Science
Female	8	6	11	5
Male	10	4	8	4

(b) What is the percentage of female students who like English as their favorite subject?

(c) A person looking at this table concludes that it is more likely that a female student will like social studies than a male student will like math. Is this correct? Justify your answer.

(d)Is it more likely that a person who likes social studies will be female or that a person who is female will like social studies? Justify.

5.) Demographers are trying to understand the association between where a person lives and how they commute to work. They survey 100 people in three cities with the results shown below.

(a) What is the frequency given that a person rides a train to work in New York?

	Car	Train	Walk
New York	5	25	10
Los Angeles	18	12	5
Chicago	8	14	3
Total	31	51	18

(b) What is the frequency that a person who lives in Los Angeles drives a car to work?

(c) What is the percentage of people walking to work in Chicago?

(d) Is a person more likely to ride a train if they live in New York or if they live in Chicago? Justify your answer.

____ 6.) When $9x^2 - 100$ is factored, it is equivalent to $(3x - b)(3x + b)$. What is the value of b?

[1] 50   [3] 3

[2] 10   [4] 100

____ 7.) What is the slope of a line passing through points $(-7,5)$ and $(5, -3)$?

[1] $-\frac{3}{2}$ [3] $\frac{2}{3}$

[2] $-\frac{2}{3}$ [4] $\frac{3}{2}$

8.) On the set of axes below, solve the following system of inequalities graphically. State the coordinates of a point in the solution set.

$y + 3 < 2x$ $-2y \leq 6x - 10$
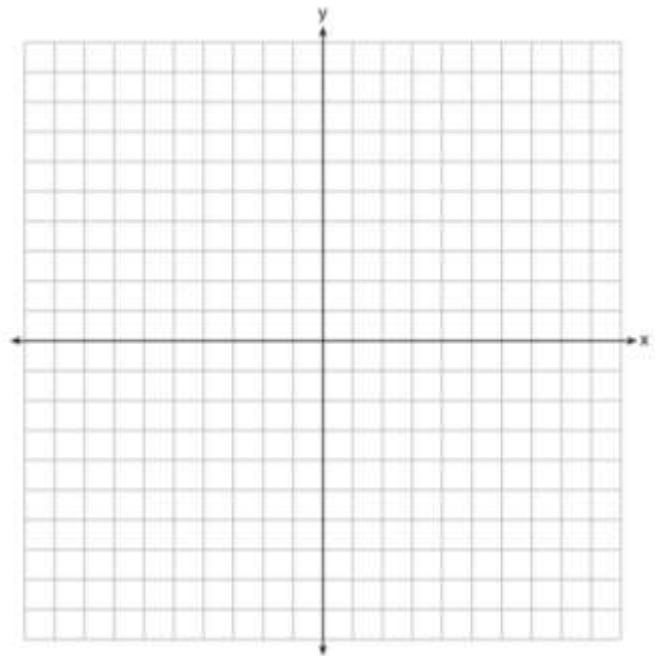
Name: _____ ^{*Homework Answers*}

Date: _____ Period: _____
Algebra   Two Way Frequency Tables   14E HW

1.) [4]

2.) 33 %

3.) a.) $\frac{29}{74}$ b.) 41 %   c.) 15%

4.) a.) $\frac{4}{13}$ b.) 20%   c.) No   d.) It is more likely for a person who likes social studies to be female.

5.) a.) $\frac{25}{51}$ b.) $\frac{18}{35}$ c.) 17%   d.) It is more likely for a person to ride the train in New York.

6.) [2]

7.) [2]

8.) Graph (5,1)

Name: _____   Date: _____ Period: _____
Algebra   Linear Regression on Calculator   14F

Exercise #1: A survey was taken of 8 low and high temperatures, in Fahrenheit, in the month of April to try to establish a relationship between a day's low temperature and high temperatures.

Low Temperature, x	22	24	26	28	30	32	34	36
High Temperature, y	30	38	34	42	53	46	58	56

(a) Given the following data points, plot on the following graph grid.
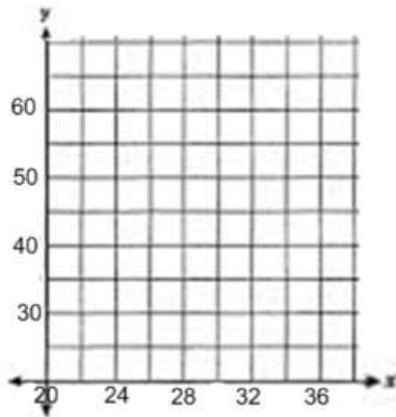
This is called a **Scatter Plot**.
Scatter plots are similar to line graphs in that they use horizontal and vertical axes to plot data points. However, they have a very specific purpose. Scatter plots show how much one variable is affected by another. The relationship between two variables is called their **correlation** .

The **line of best fit** , also called a trendline or a **linear regression**, is a straight line that best illustrates the overall picture of what the collected data is showing. It helps us to see if there is a relationship or correlation between the two factors being studied. This trendline helps us to predict future events relating to the data being studied.
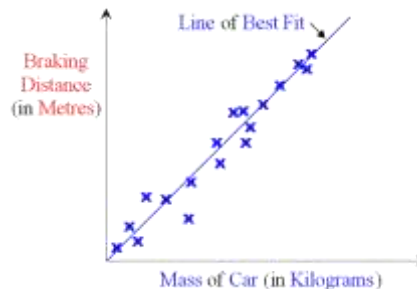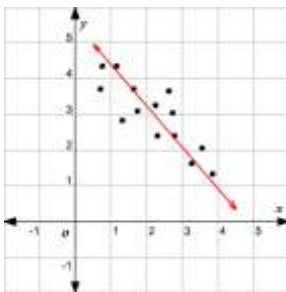
Exercise #2: Which scatter plot has the most accurate line of best fit?

(1)

(2)

(3)

(4)
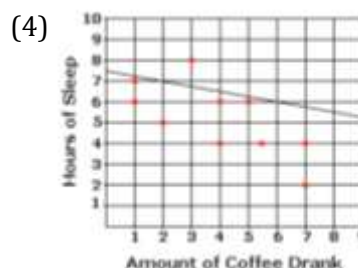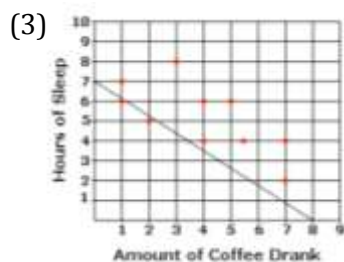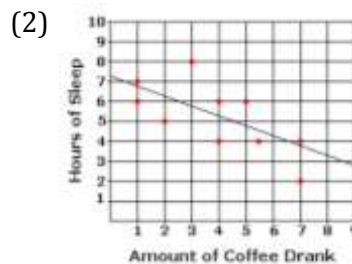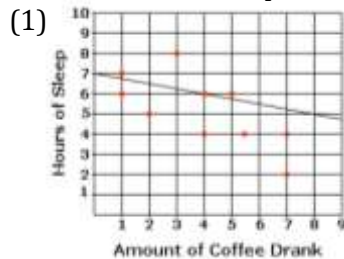
We have now discovered how to sketch the line of best fit. However, this isn't 100% accurate. Now we are going to learn how to get the line of best fit using our calculator.

Let's take a closer look at **Exercise #1:**

Exercise #3: A survey was taken of 8 low and high temperatures, in Fahrenheit, in the month of April to try to establish a relationship between a day's low temperature and high temperatures.

Low Temperature, x	22	24	26	28	30	32	34	36
High Temperature, y	30	38	34	42	53	46	58	56

(a) Use your calculation to find the equation for the line of best fit. Round the slope of the line to the nearest hundredth and the y-intercept to the nearest integer.

How to find the Equation of the Line of Best Fit (Linear Regression):
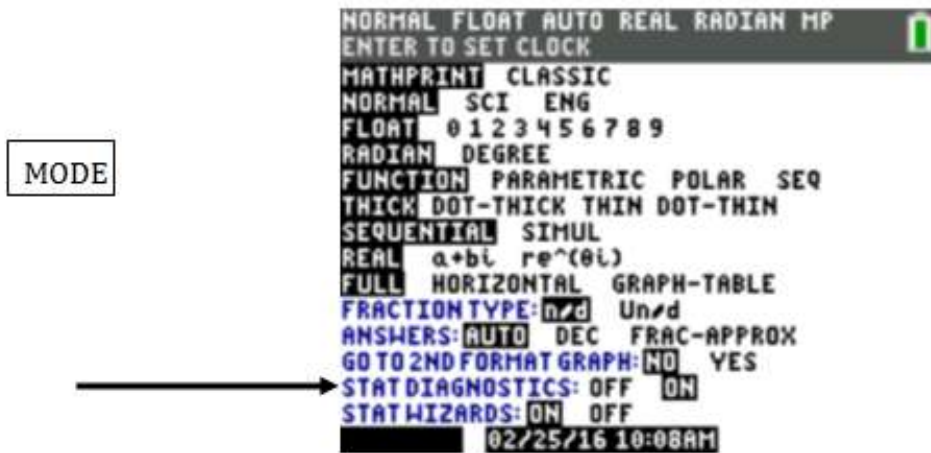
Step 1: STAT → 1: Edit....

Step 2: Insert information into two lists:
$L_1 = x$ values   **Make sure both lists have the same amount**
$L_2 = y$ values

Step 3: STAT → CALC → 4: LinReg (ax+b)

Step 4: Make sure your Xlist and Ylist are L_1 and L_2. Then, scroll down to **Calculate** and hit **Enter**

Step 5: Use given information to create the Line of Best Fit.

Equation for the Line of Best Fit:

Exercise #4: Generally, the fuel efficiency of a car changes with the weight of the car. A survey of some cars with their weights and gas mileages is shown below.

Weight (1000's of lbs)	3.7	4.5	3.2	5.1	6.8	4.9	4.8	5.5
Mileage (miles per gallon)	38	26	48	24	18	30	28	21

(a) Find the equation for the line of best fit using your calculator. Round both coefficients to the nearest tenth. List what the variables x and y represent in this problem.

(b) Given the graph of this scatter plot, would you consider the **correlation** between weight and mileage to be **positive** or **negative?** Explain.
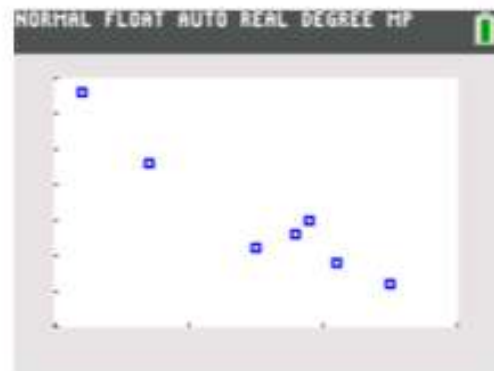
(c) Which **parameter** of the linear model predicts whether the **correlation** is positive or negative? Use this model to help explain your answer.

(d) If a car had a weight of 4,300 pounds, what would this model predict as its fuel efficiency? Round to the nearest integer. Use appropriate units and make sense of your answer.

(e) If we wanted to purchase a car that got 40 miles to a gallon, what weight of car, to the nearest 100 pounds, should we purchase? Solve algebraically.

Name: _____   Date: _____ Period: _____
Algebra   Linear Regression on Calculator   14F HW
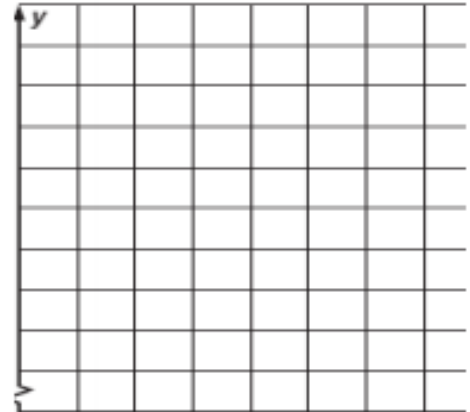
1) The table below gives the number of hours spent studying for a science exam and the final exam grade.

    (a) Draw a scatter plot of the data and sketch in the line of best fit.

    (b) What is the equation for the line of best fit? (Use calc)
         Round all values to nearest tenth.

Study Hours	3	2	5	1	0	4	3
Grade	84	77	92	70	60	90	75

    (c) Could this line go on forever? Why or why not?

2) The table shows the average and maximum longevity of various animals in captivity.

    (a) Draw a scatter plot and determine, what relationship, if any, exists.

    (b) Sketch in the line of best fit.

    (c) What is the equation for the line of best fit? (Use calc)
         Round all values to nearest tenth.

Longevity (years)								
Avg.	12	25	15	8	35	40	41	20
Max.	47	50	40	20	70	77	61	54

Animal Longevity (Years)
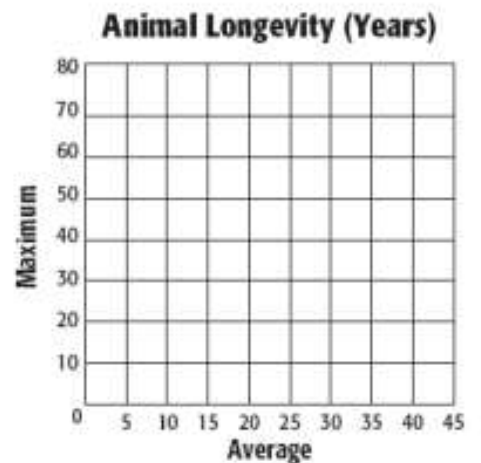
3) A survey was done at Ketcham High School to determine the effect of time spent on studying and grade point average. The table below shows the results for 10 students randomly selected.

Study time (Hours per week)	2	4	5	7	10	12	14	17	19	20
GPA (out of 100)	64	71	69	74	81	86	84	94	91	96

(a) Enter the data in your calculator and use it to generate the equation for the line of best fit. Round your slope to the nearest tenth and round your *y*-intercept to the nearest integer.

(b) According to the linear regression model from part (a), what GPA, to the nearest integer, would result from studying for 15 hours in a given week? Justify your answer.

(c) A passing average is defined as a 65% or above. Does the model predict a passing average if the student spends no time studying in a given week? Justify your answer.

(d) For each additional hour that a student studies per week, how many points does the model predict a GPA will rise? Explain how you arrived at your answer.

4) The mean annual temperature of a location generally depends on its elevation above sea level. A collection of nine locations in Nevada were chosen and had their elevation and mean annual temperature recorded. The data is shown below.

Elevation (feet)	1200	4125	6230	2378	5625	6328	4375	1864	3160
Mean Temperature (°F)	62	45	36	51	48	32	40	58	49

(a) Use your calculator to determine the equation for the line of best fit. Round your slope to the nearest *thousandth*. Note that it will be a small number. Round your *y*-intercept to the nearest integer.

(b) What does the *y*-intercept tell you about the temperature in Nevada?

(c) Using correct units, give an interpretation of the slope of this line.

(d) Using your model from part (a), what would be the predicted mean temperature at an elevation of 3000 feet above sea level?

(e) Would you characterize this correlation as being positive or negative? How can you tell this from the equation itself?

5) In the function $f(x) = (x + 1)^2 + 4$, identify the vertex.

____ 6) When solving the equation $x^2 - 8x - 7 = 0$ by completing the square, which equation is a step in the process? ==**If you don't have work you will not receive full credit on your homework**==

(1) $(x - 4)^2 = 9$ (3) $(x - 4)^2 = 23$

(2) $(x - 8)^2 = 9$ (4) $(x - 8)^2 = 23$

Name: ___*Homework Answers*_____   Date: _____ Period: _____
Algebra   Linear Regression on Calculator   14F HW

1.) (a) Sketch a line of best fit   (b) $y = 6.3x + 62.0$
    (c) No; with appropriate explanation

2.) (a) Positive Correlation   (b) Sketch a line of best fit
    (c) $y = 1.3x + 22.4$

3.) (a) $y = 1.7x + 62$ (b) 88
    (c) No; with appropriate explanation   (d) GPA will rise 1.7 points for each additional hour of studying.

4.) (a) $y = -.005x + 65$ (b) The mean temperature is 65 degrees Fahrenheit.
    (c) For every foot of increase above sea level, the mean temperature decreases by .005 degrees Fahrenheit.
    (d) 50 degrees Fahrenheit   (e) Negative correlation; the slope is negative.

5.) (-1,4)

6.) (3)

Name: _____   Date: _____ Period: _____
Algebra   Other Types of Regression   14G

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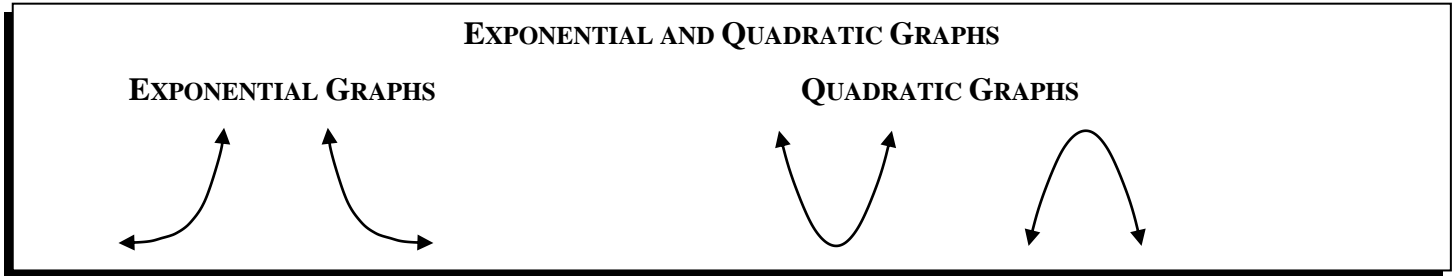
***Exercise* #1:** For each scatterplot shown below, determine if it is best fit with a linear, exponential, or quadratic function. Draw a curve of best fit depending on your choice.

(a)

Type: _____

(b)

Type: _____

(c)

Type: _____

(d)

Type: _____

(e)

Type: _____

(f)
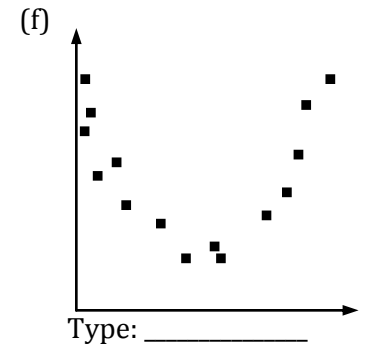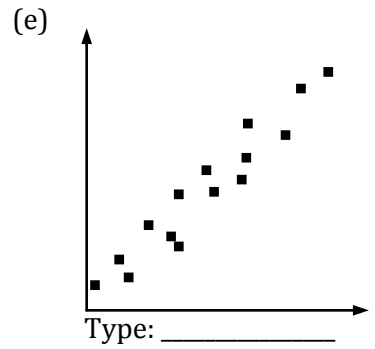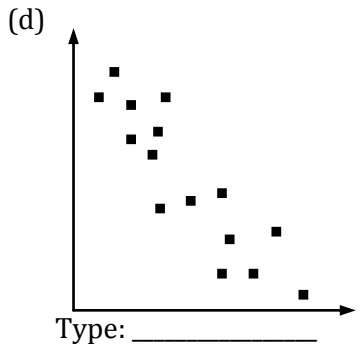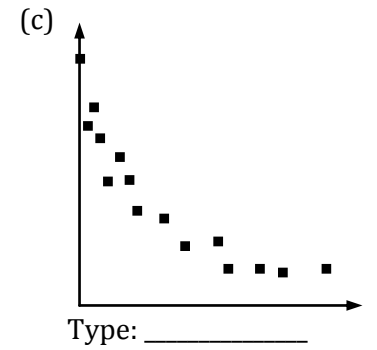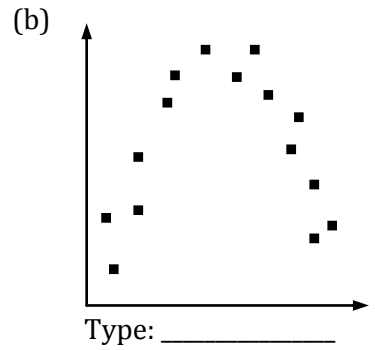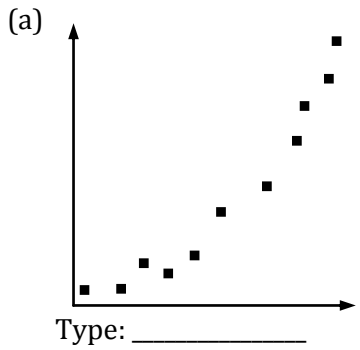
Type: _____

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*.

==**To find the exponential regression equation, insert lists into L_1 and L_2. Click on STAT, CALC. Instead of using 4:LinReg, use 0: ExpReg**==

(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?

Exercise #3: An application developer released a new app to be downloaded. The table below gives the number of downloads for the first four weeks after the launch of the app.

| Number of Weeks	1	2	3	4
Number of Downloads	120	180	270	405

(a) Write an exponential equation that models this data.

(b) Use this model to predict how many downloads the developer would expect in the 26th week if this trend continues. Round your answer to the *nearest download*.
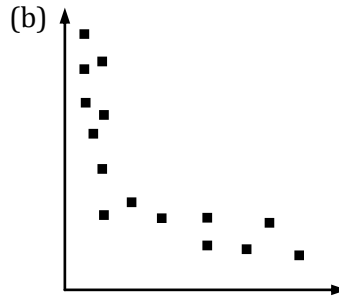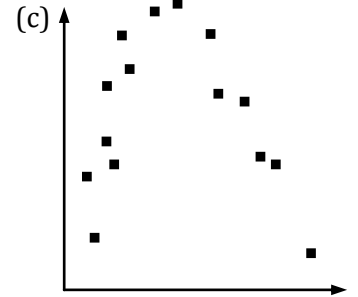
1. For each scatterplot below, determine the best type of regression from: linear, exponential, or quadratic. Draw a representative curve (line, exponential, or parabola) through the data.
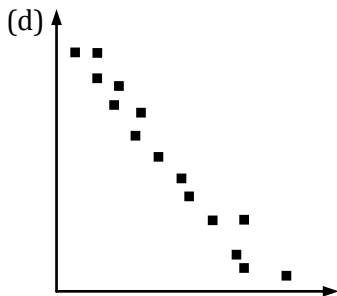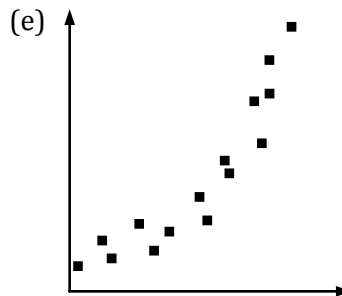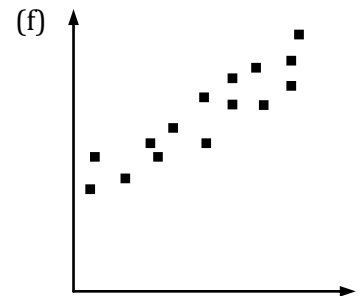
(a)

Type: _____
_____

(b)

Type_____

(c)

Type:

(d)

Type: _____
_____

(e)

Type: _____

(f)
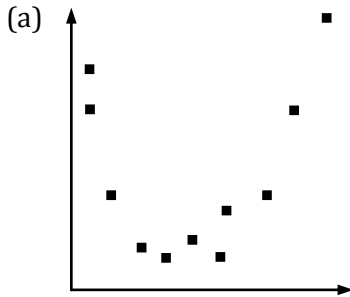
Type:

____ 2. Given the scatterplot below, which of the following equations would best model the data?
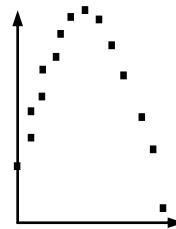Explain your choice.
[1] $y = -3x + 6$
[2] $y = 6(2)^x$
[3] $y = -4x^2 + 20x + 3$
[4] $y = 2x^2 - 6x + 4$

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

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

4. 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

Days After the Website Launched
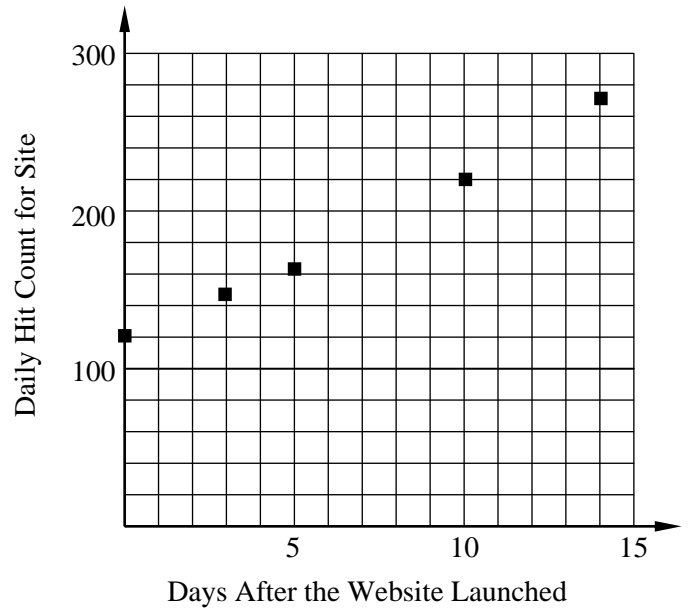
(a) Of the two (Linear/Exponential) types of regression we have studied which seems most likely to fit this data? Explain your choice.

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

5. The cost of belonging to a gym can be modeled by $C(m) = 50m + 79.50$, where $C(m)$ is the total cost for m months of membership.

State the meaning of the slope and y-intercept of this function with respect to costs associated with the gym membership.

6. When $(2x - 3)^2$ is subtracted from $5x^2$, the result is _____?

Name: _____ *Homework Answers*   Date: _____

Period: _____

Algebra   Other Types of Regression

14G HW

1.) (a) Quadratic   (b) Exponential   (c) Quadratic
     (d) Linear   (e) Exponential   (f) Linear

2.) (3)

3.) The outputs decrease and the increase. Linear/Exponential only increase or only decrease.

4.) (a) Linear; with appropriate explanation
     (b) Linear: $y = 10.83x + 114.07$ Exponential: $y = 121.09(1.06)^x$
     (c) Exponential; grows faster over time

5.) Proper explanations

6.) $x^2 + 12x - 9$