

Unit 6 Check Sheet

Name _____ Per _____

Data Analysis

(Print)

- Check sheet must be turned in to receive Homework & Quiz points.
- All quiz corrections must be done for test score to replace quiz scores.
- No check sheet = No Points.
- Write quiz scores as fractions
- Lost Quizzes count as a 0.
- Quiz ratio is total points scored on quizzes and pre-test out of total possible
- Order (from top to bottom)
 - Check sheet,
 - **Quiz 1, 2, Pre-Test**
 - **Quiz corrections**

Section	HMK
6.1 Frequency and Histograms Worksheet 6.1 #1-15 all	
6.2 Measures of Central Tendency and Dispersion Worksheet 6.2 #1-11 all	
6.3 Box-and-Whisker Plots Worksheet 6.3 #1-14 all Matching Flashcards Activity Quiz 1	
6.4 Scatter Plots and Trend Lines Worksheet 6.4 #1-11 all	
6.4B Linear Regression Worksheet 6.4B #1-5 all	
6.4C Residuals Worksheet 6.4C #1-5 all	
6.5 Two-Way Frequency Tables Worksheet 6.5 #1-19 all Flashcard Activity Quiz 2	
Review Review Worksheet #1-11 all Unit 6 Pre-Assessment	
Unit Test	

Quiz 1: _____ Score/Possible

Quiz 2: _____ Score/Possible

Pre-Test: _____ Score/Possible

Total Quiz Ratio: _____ Total Score/Total Possible

6.1 Practice

Form G

Frequency and Histograms

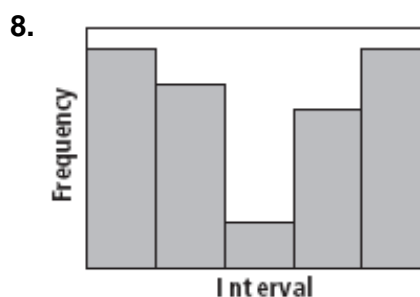
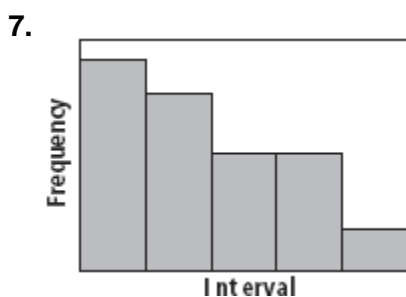
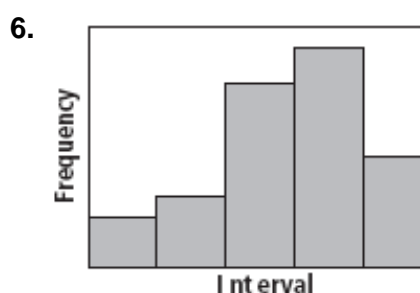
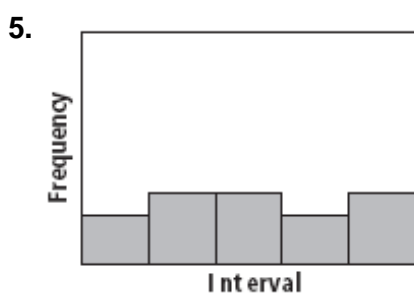
Use the data to make a frequency table.

1. runs per game: 5 4 3 6 1 9 3 4 2 2 0 7 5 1 6
2. weight (lb): 10 12 6 15 21 11 12 9 11 8 8 13 10 17

Use the data to make a histogram.

3. number of pages: 452 409 355 378 390 367 375 514 389 438 311 411 376
4. price per yard: \$9 \$5 \$6 \$4 \$8 \$9 \$12 \$7 \$10 \$4 \$5 \$6 \$6 \$7

Tell whether each histogram is *uniform*, *symmetric*, or *skewed*.



Practice (continued)

Form G

Frequency and Histograms

Use the data to make a cumulative frequency table.

9. call length (min): 3 5 12 39 12 3 15 23 124 2 1 1 7 19 11 6

10. package weight (kg): 1.25 3.78 2.2 12.78 3.15 4.98 3.45 9.1 1.39

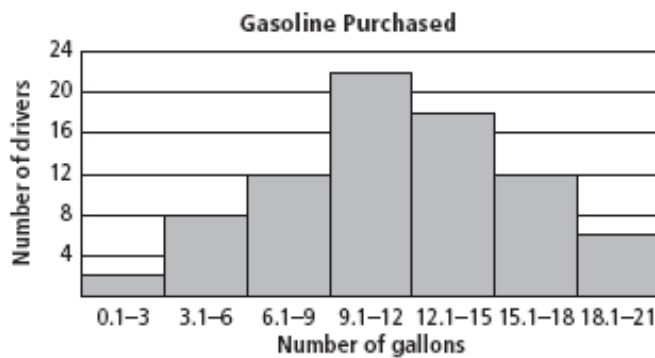
Use the snowfall amounts, in inches, below.

10 2.5 1.5 3 6 8.5 9 12 2 0.5 1 3.25 5 6.5 10.5 4.5 8 8.5

11. What is a histogram of the data that uses intervals of 2?

12. What is a histogram of the data that uses intervals of 4?

The amount of gasoline that 80 drivers bought to fill their cars' gas tanks is shown.



13. Which interval represents the greatest number of drivers?

14. How many drivers bought more than 12 gallons?

15. How many drivers bought 9 gallons or less?

Name _____ Period _____ Date _____

Math 1 Worksheet 6.2: Central Tendency and Dispersion

Find the mean, median and mode of each data set. Explain which measure of central tendency best describes the data

1. Price per item:

\$4 \$14 \$15 \$20 \$15

2. Extra points kicked:

3 5 5 2 4

3. Average speed (mi/hr):

36 59 47 56 67

4. Price per pound:

\$30 \$8 \$2 \$5 \$6

Find the value of x such that the data set has the given mean.

5. 14, 10, 17, 9, x ; mean of 14

6. 55, 60, 35, 90, x ; mean of 51

7. 6.5, 4.3, 9.8, 2.2, x ; mean of 4.8

8. 0.9, 1.6, 3.4, 0.5, x ; mean of 1.4

9. Anne's golf scores for the season are 88, 90, 86, 89, 96, and 85. Chi's golf scores are 91, 86, 88, 84, 90, and 83. Find the range and mean of each golfer. Based on the data, explain who you think is the better golfer (lower scores are better in golf).

10. Armand and Tyler are two of your employees. You have looked at their work history to find the average number of orders that they fill each day for the past four weeks:

Armand: 22.4, 20, 33.5, 21.3

Tyler: 6.2, 15, 50.4, 28

Use the mean and range to compare their data. Who is the better worker? Why?

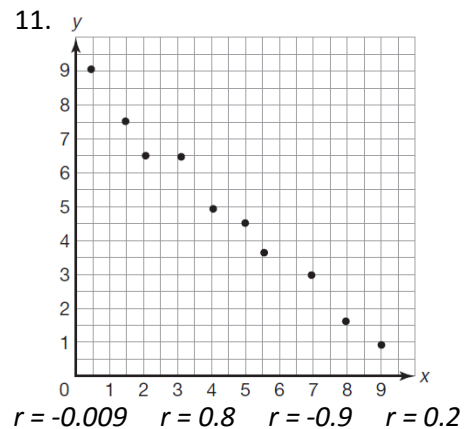
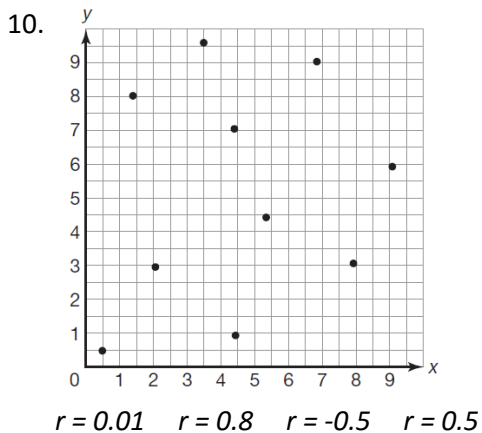
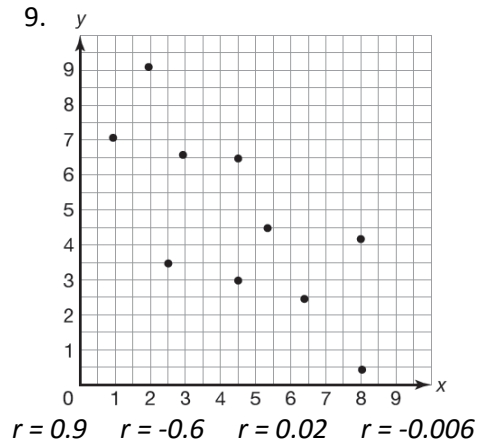
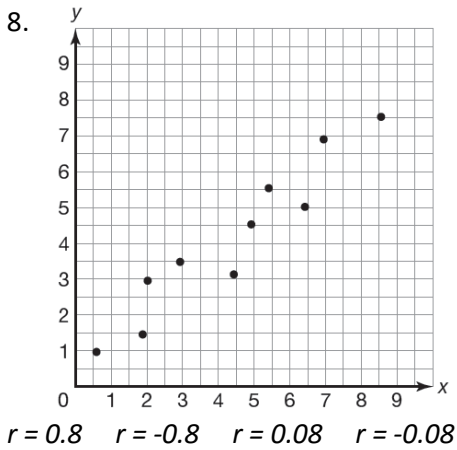
11. A math teacher was looking at the most recent quiz scores for one of his classes:

5 6 10 9 9 10 5 10 6 9 8 8 7 10 9 8 5 7 7 9 9 8 10 8 9

Create a line plot of the data. Is the data symmetrical, skewed, or uniform? If you found the mean and the median, which one would be greater? Why?

6. During one month at a local deli, the amount of ham sold decreased as the amount of turkey sold increased. Is this an example of positive correlation, negative correlation, or no correlation?
7. In each of the following situations, is there likely to be a correlation? If so, does the correlation imply causation?
- The cost of a family's vacation and the size of their house.
 - The time spent exercising and the number of Calories burned.
 - The amount of time you study for a test and the score you receive.
 - A person's height and the number of letters in the person's name.

Determine whether the points in each scatter plot have a positive correlation, a negative correlation, or no correlation. Then estimate a correlation coefficient that is most accurate.



Name _____ Period _____ Date _____

Math 1 Worksheet 6.4B: Linear Regression

Use a graphing utility to find the linear regression equation for the give data.

1. A student who waits on tables at a restaurant recorded the cost of meals and the tip left by single diners.

Meal cost	\$4.75	\$6.84	\$12.52	\$20.42	\$8.97
Tip	\$0.50	\$0.90	\$1.50	\$3.00	\$1.00

If the next diner orders a meal costing \$10.50, how much tip should the waiter expect to receive?

Equation: _____ Tip expected: _____

2. The table below gives the number of hours spent studying for a science exam (x) and the final exam grade (y).

x	2	5	1	0	4	2	3
y	77	92	70	63	90	75	84

Predict the exam grade of a student who studied for 6 hours.

Equation: _____ Grade expected: _____

3. The table below shows the lengths and corresponding ideal weights of sand sharks.

Length	60	62	64	66	68	70	72
Weight	105	114	124	131	139	149	158

Predict the weight of a sand shark whose length is 75 inches.

Equation: _____ Weight expected: _____

4. The table below gives the height and shoe sizes of six randomly selected men.

Height	67	70	73.5	75	78	66
Shoe size	8.5	9.5	11	12	13	8

If a man has a shoe size of 10.5, what would be his predicted height?

Equation: _____ Height expected: _____

5. A convenience store manager notices that sales of soft drinks are higher on hotter days, so he assembles the data in the table.

High temperature (*F)	55	58	64	68	70	75	80	84
# cans sold	340	335	410	460	450	610	735	780

Predict soft-drink sales if the temperature is 95°F.

Equation: _____ Sales expected: _____

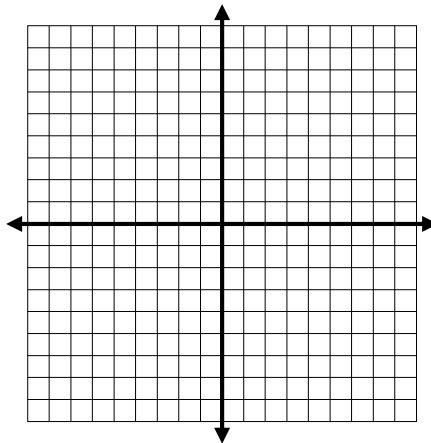
Math 1 Worksheet 6.4C: Residuals

Complete the table of residual values and construct a residual plot.

1. Linear regression equation: $y = 0.5x$

x	y	Predicted Value	Residual Value
5	3		
10	4		
15	9		
20	7		
25	13		
30	15		

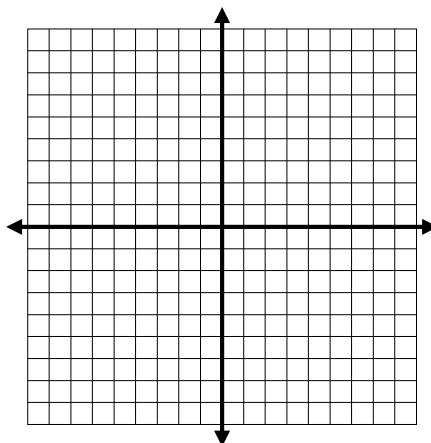
Residual plot:



2. Linear regression equation: $y = -0.4x + 16.3$

x	y	Predicted Value	Residual Value
2	5		
4	15		
6	26		
8	23		
10	11		
12	3		

Residual plot:

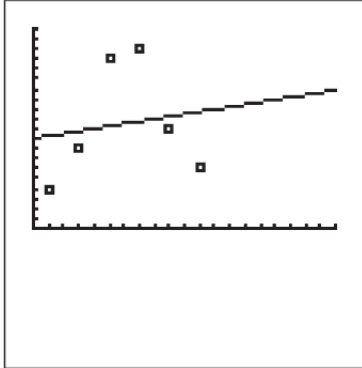


Consider the scatter plot, its line of best fit, and the corresponding residual plot of each data set. State if a linear model is appropriate for the data and explain why or why not.

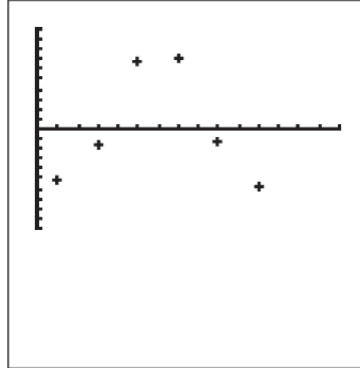
3. Linear regression equation: $y = 0.24x + 9.04$; $r = 0.1570$

x	1	3	5	7	9	11
y	4	8	17	18	10	6

Scatter Plot & Line of Best Fit:



Residual Plot:

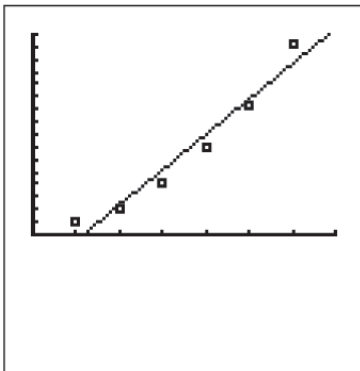


Is a linear model a good fit?

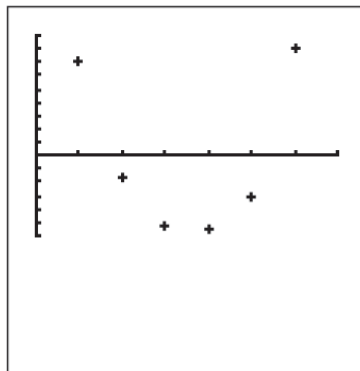
4. Linear regression equation: $y = 14.08 - 163.13x$; $r = 0.9746$

x	10	20	30	40	50	60
y	49	103	207	346	511	762

Scatter Plot & Line of Best Fit:



Residual Plot:

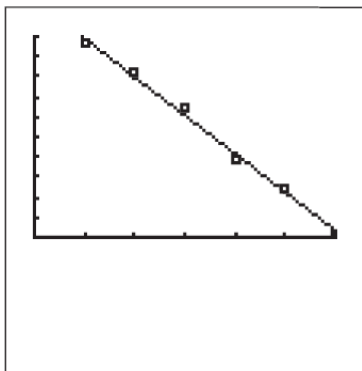


Is a linear model a good fit?

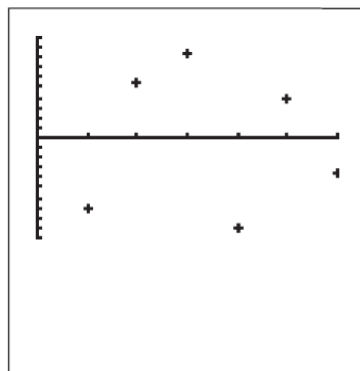
5. Linear regression equation: $y = -1.91x + 59$; $r = -0.9968$

x	5	10	15	20	25	30
y	48	41	32	19	12	1

Scatter Plot & Line of Best Fit:



Residual Plot:



Is a linear model a good fit?

Name _____ Period _____ Date _____

Math 1 Worksheet 6.3: Box and Whisker Plots

Use the following data set for problems 1–8.

The class scores for the Unit 6 Test are given below:

88, 76, 77, 95, 69, 72, 83, 88, 99, 83, 75, 62, 45, 76, 93, 84, 81, 74, 68

1. Sort your data from smallest to largest
2. Find the minimum, Q1, median, Q3, and maximum
3. Find the IQR
4. Create a box-and-whisker plot of the data



5. Identify any outliers in the data set
6. Create a box-and-whisker plot for the data **without** the outlier



7. Compare the two box-and-whisker plots. How does the outlier affect the graph?
8. What affect will the outlier have on the mean of the data?

For problems 9-10, create box and whisker plots for the two data sets (one above the line and one below) and then compare the results

9. Weekly cars sold:

Kathy's: 5 8 3 12 7 11 9 4 8

Samuel's: 9 2 5 10 9 7 7 6 10



10. Video length (min):

Training 1: 78 62 45 65 50 59 67 62 51 70

Training 2: 60 67 50 58 62 71 69 54 60 64



11. Of 200 Math 1 Unit 5 test scores, 32 are less than or equal to 60.

What is the percentile rank of a score of 60?

12. Of 25 dogs, 15 weigh more than 35 pounds. What is the percentile rank of a dog that weighs 35 pounds?

13. The class of 2015 had 631 students who graduated. 481 students had a GPA of 3.5 or less. What is the percentile rank of a person who had a GPA of 3.5?

6.5 Practice

Form G

Two-Way Frequency Tables

Weekend Mornings The table shows the results of a survey asking students whether they do or do not play a sport and whether they prefer to exercise or sleep in on weekend mornings.

Plays a Sport	Preferred Way to Spend Weekend Mornings		
	Exercise	Sleep In	Total
Yes	40	70	110
No	55	85	140
Total	95	155	250

Use the table to answer Exercises 1–12.

1. What type of table is shown?
2. Identify the variables and categories for the data.
3. How many students took part in the survey?
4. What is the joint frequency of students who play a sport and prefer to exercise on weekend mornings?
5. What is the joint frequency of students who don't play a sport and prefer to sleep in on weekend mornings?
6. What is the marginal frequency of students who don't play a sport?
7. What is the marginal frequency of students who prefer to exercise weekend mornings?
8. What is the joint relative frequency of students who play a sport and prefer to sleep in weekend mornings?
9. What is the marginal relative frequency of students who prefer to exercise weekend mornings?
10. What is the conditional relative frequency that a student does not play a sport, given that the student prefers to exercise weekend mornings?
11. What is the conditional relative frequency that a student plays a sport, given that the student prefers to exercise weekend mornings?
12. **Reasoning** What conclusions can you draw based on the conditional relative frequencies you found in Exercises 10 and 11?

Practice (continued)

Form G

Two-Way Frequency Tables

Use the following to answer Exercises 13–19.

Favorite Music A survey asked 325 males and females whether their favorite music was country, hip-hop, or rock. Of the 143 males, 26 liked country best and 39 preferred rock. Of the 182 females, 52 named hip-hop as their favorite. A total of 91 people said that country was their favorite.

13. Make a two-way frequency table using the given information. Then find the missing frequencies to complete the table.

14. Make a two-way relative frequency table for the data.

15. What is the joint relative frequency of males who like country music best?

16. What is the marginal relative frequency of males surveyed?

17. What is the conditional relative frequency that a person surveyed was female, given that the person liked country best?

18. What is the conditional relative frequency that a person surveyed was female, given that the person liked hip-hop best?

19. What is the conditional relative frequency that a person surveyed was female, given that the person liked rock best?

*6.1-6.3

Frequency

* 6.1-6.3

Frequency
Table

*6.1-6.3

Histogram

*6.1-6.3

Cumulative
Frequency
Table

*6.1-6.3

Uniform
Histogram

*6.1-6.3

Symmetrical
Histogram

*6.1-6.3

Skewed
Histogram

*6.1-6.3

3 Measures
of Central
Tendency

*6.1-6.3

Outlier

*6.1-6.3

Mean

*6.1-6.3

Median

*6.1-6.3

Mode

*6.1-6.3

Measures of
Dispersion

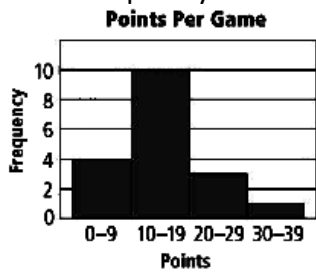
*6.1-6.3

Range

*6.1-6.3

Line Plot

6.1-6.3 graph that displays data from a frequency table



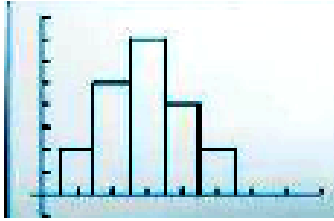
6.1-6.3 groups a set of data values into intervals and shows the frequency of each interval

Home Runs	Frequency
2-6	4
7-11	5
12-16	4
17-21	1

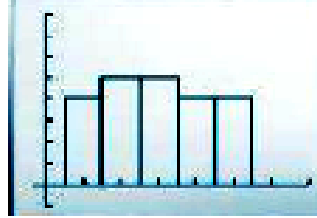
6.1-6.3

The number of data values

6.1-6.3 A vertical line can divide the histogram into two parts that are close to mirror image



6.1-6.3 Histogram with bars roughly the same height



6.1-6.3 shows number of data values that lie in or below a given interval

Interval	Frequency	Cumulative Frequency
0-4	6	6
5-9	4	10
10-14	4	14
15-19	2	16

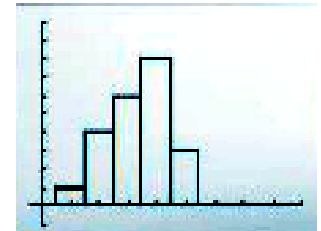
6.1-6.3

A data value that is much greater than OR less than the other values in the set.

6.1-6.3

Mean, Median, Mode

6.1-6.3 Histogram with one peak that is not in the center



6.1-6.3

Data value that occurs the most times. Best used when data is NOT numbers or when choosing the most popular item.

6.1-6.3

Middle value in a data set when arranged in order OR average of two middle values. Best used when data **does** have an outlier.

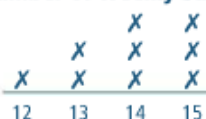
6.1-6.3

Sum of numbers divided by the number of data values. Best used when data **does not** have an outlier.

6.1-6.3

Data display in which each mark above a number line corresponds to each data value. Also called dot plots.

Number of Weekly Sales



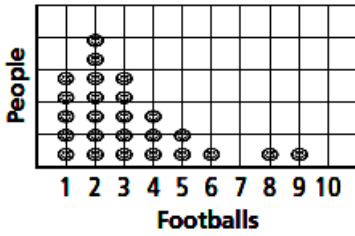
6.1-6.3

Difference between the greatest and least data value.

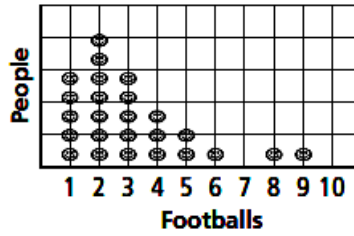
6.1-6.3

Describe how spread out the values in a data set are. Ex. range

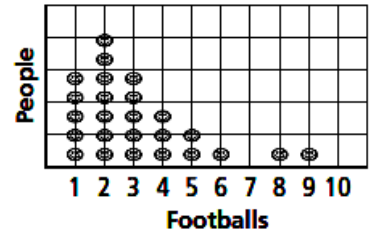
*6.1-6.3 What is the mean of the line plot?



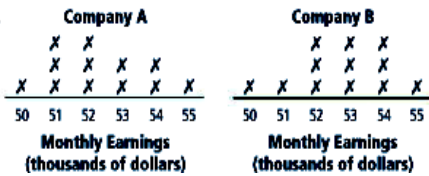
* 6.1-6.3 What is the mode of the line plot?



*6.1-6.3 What is the median of the line plot?



*6.1-6.3 Which has a greater mean and/or median and why?



*6.1-6.3 What is the definition of standard deviation?

*6.1-6.3 How do you compute the standard deviation of:

4, 8, 10, 12, 16

*6.1-6.3

Box and Whisker Plot

*6.1-6.3

Quartiles
Q₁ Q₂ Q₃

*6.1-6.3

Minimum

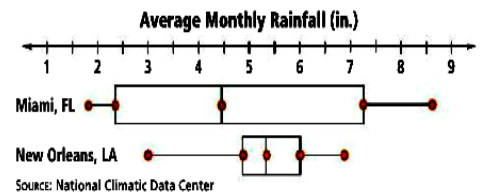
*6.1-6.3

Maximum

*6.1-6.3

Interquartile Range

*6.1-6.3 State the similarities and differences.



*6.4-6.5

Scatter Plot

*6.4-6.5

Positive Correlation

*6.4-6.5

Negative Correlation

6.1-6.3

3

because it is the middle value

6.1-6.3

2

because it is the value that occurs the most

6.1-6.3

3.16

because it is the sum of values divided by how many numbers

6.1-6.3

1. Find the mean = 10
2. Subtract each data point from the mean
3. Square each value
4. Find the ave of "squares" = 16
5. Take square root of ave = 4

6.1-6.3

Measure of dispersion that measures how the values in a data set vary from the mean

6.1-6.3

B: greater mean
greater median

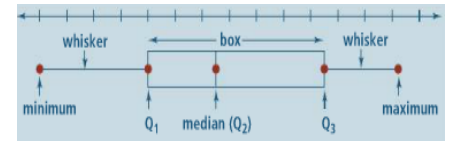
6.1-6.3

Smallest data value

6.1-6.3

Q_1 = median of lower half of data
 Q_2 = median of all data
 Q_3 = median of upper half of data

6.1-6.3 Graph that summarizes the data



6.1-6.3

Miami: 50% of rainfall varies more widely = larger interquartile range
New Orleans: Range of 4 is smaller than Miami's range of 7. Has higher median value than Miami.

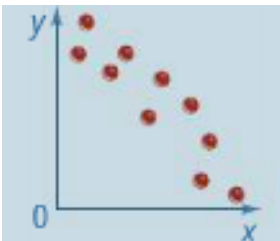
6.1-6.3

The difference between the third and first quartile. $Q_3 - Q_1$

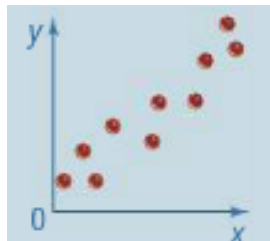
6.1-6.3

Largest data value

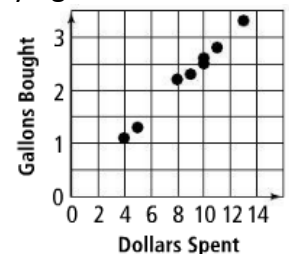
6.4-6.5 As x increases, the y decreases



6.4-6.5 As x increase, the y increases



6.4-6.5 Graph that relates two different sets of data by displaying them as ordered pairs



*6.4-6.5

No
Correlation

*6.4-6.5

Trend Line

*6.4-6.5

Line of
Best Fit

*6.4-6.5

Correlation
Coefficient

*6.4-6.5

Correlation
Coefficient of
 $r = -1$

*6.4-6.5

Correlation
Coefficient of
 $r = 0$

*6.4-6.5

Correlation
Coefficient of
 $r = 1$

*6.4-6.5

Correlation
Coefficient of
 $r = -0.8$

*6.4-6.5

Correlation
Coefficient of
 $r = 0.8$

*6.4-6.5

Causation

*6.4-6.5

Joint
Frequency

*6.4-6.5

Marginal
Frequency

*6.4-6.5

Joint
Relative
Frequency

*6.4-6.5

Marginal
Relative
Frequency

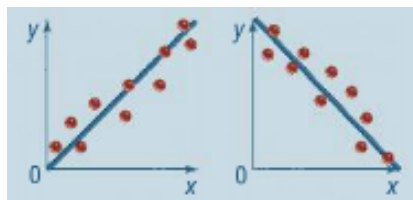
*6.4-6.5

Conditional
Relative
Frequency

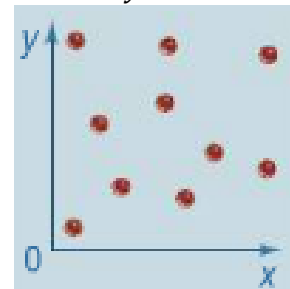
6.4-6.5

The BEST trend line on a scatter plot drawn near the points.

6.4-6.5 A line on a scatter plot drawn near the points.



6.4-6.5 x and y are not related.



6.4-6.5

No Correlation

6.4-6.5

strong negative correlation

6.4-6.5

Number from -1 to 1 that tells you how closely the equation models the data. (r)

6.4-6.5

medium positive correlation

6.4-6.5

medium negative correlation

6.4-6.5

strong positive correlation

6.4-6.5 an entry in the total row or column

Grade	Preferred Subject		
	Math	English	Total
Freshman	36	51	87
Sophomore	56	27	83
Total	92	78	170

6.4-6.5 an entry in the body of the table

Grade	Preferred Subject		
	Math	English	Total
Freshman	36	51	87
Sophomore	56	27	83
Total	92	78	170

6.4-6.5

When a change in one quantity causes a change in a second quantity. Correlation does not always imply causation.

6.4-6.5 Divide a joint frequency by that frequency's row total or column total

Grade	Preferred Subject		
	Math	English	Total
Freshman	36	51	87
Sophomore	56	27	83
Total	92	78	170

Prefers math, given freshman = $\frac{36}{87}$

Soph, given prefers math = $\frac{56}{92}$

6.4-6.5 Divide a row total or a column total by the grand total

Grade	Preferred Subject		
	Math	English	Total
Freshman	36	51	87
Sophomore	56	27	83
Total	92	78	170

Mar. Rel. Freq of Freshman = $\frac{87}{170}$

Mar. Rel. Freq of prefer Eng = $\frac{78}{170}$

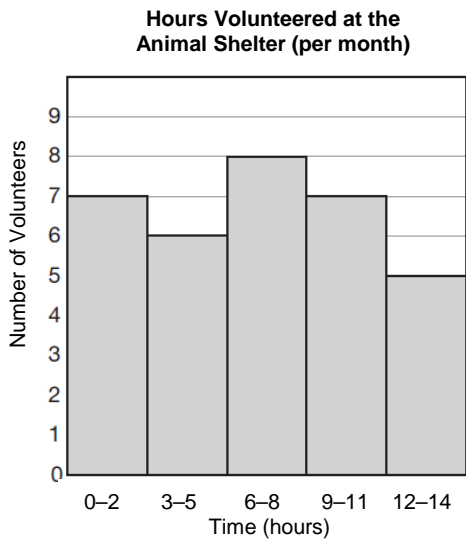
6.4-6.5 Divide a frequency in the body of the table by the grand total

Grade	Preferred Subject		
	Math	English	Total
Freshman	36	51	87
Sophomore	56	27	83
Total	92	78	170

Joint Rel. Freq of Freshmen who prefer math = $\frac{36}{170}$

Math 1 Unit 6 Review

1.



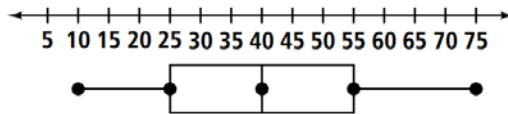
How many total people volunteered at the shelter this month?

How many volunteers worked less than 6 hours in the month?

How many volunteers worked between 3 and 8 hours in the month?

2. Which measure of central tendency (mean, median, or mode) is MOST affected by an outlier? Why?

3. For the box-and-whisker plot, identify the following:



Minimum:

Q1:

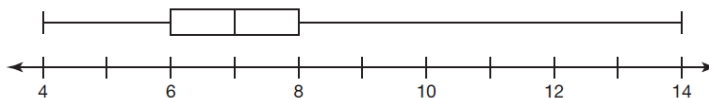
Median:

Q3:

Maximum:

IQR:

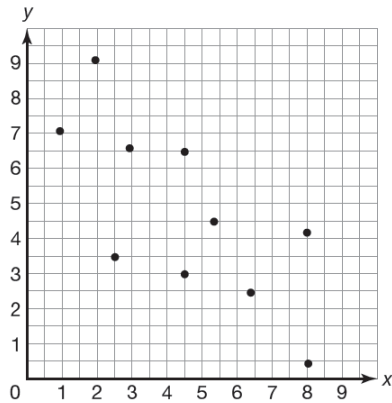
4. The box-and-whisker plot shows the hours of television 8th graders watched in a week.



What percent of students watched more than 8 hours of TV?

What is the best measure of central tendency for this data? Why?

5. Use the scatter plot shown:



What type of correlation is represented by the data?

What correlation coefficient would be accurate for a trend line for the scatter plot?

6. For the following situation, determine if there is a correlation. If there is a correlation, is it a causal relationship? Why?

the number of hours spent studying and your grade on the Unit 6 test

7. A survey was taken asking students which subject they preferred in school. The results are shown in the table below:

	Subject		
	ELA	Math	Total
Females	54	76	130
Males	78	92	170
Total	132	168	300

What is the marginal frequency of total students who prefer math?

What is the joint relative frequency of females who prefer ELA?

What is the conditional relative frequency that a student is male, given the student prefers math?

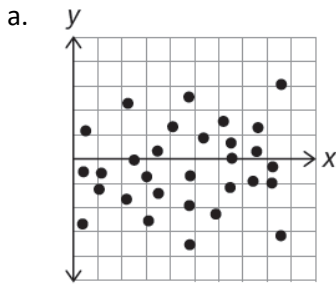
8. Jonah practices shooting 50 baskets from different distances. His data can be modeled by the equation $y = -2.5x + 42$, where x is the distance in feet and y is the number of baskets made.

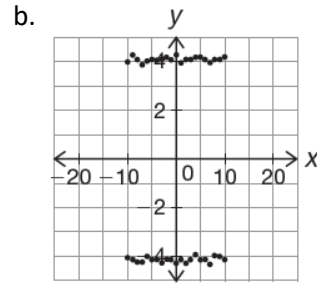
Using the equation, how many baskets do you expect Jonah to make from 7 feet?

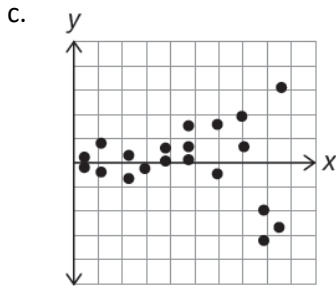
Using the equation, how many baskets do you expect Jonah to make from 15 feet?

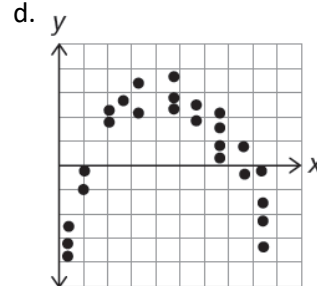
What type of correlation does this show? Why?

9. Does the residual plot suggest a linear model may be the best fit for the data? Explain why or why not.









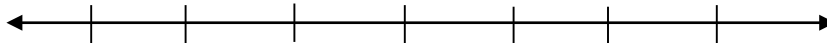
10. Complete the residual table for the linear regression equation $y = -0.3x + 8.7$

x	y	Predicted Value	Residual Value
2	6		
4	16		
6	27		
8	24		
10	12		
12	4		

11. The table shows the number of library books check out by students for the month:

Student	Jon	Tami	Mike	Karen	Sue	Steve	Doug	Patty	Chris	Charlotte	Andy	Lisa
# of books	2	3	5	1	2	7	3	4	3	4	5	6

Construct a line plot of the data:



Which measure of central tendency best describes the data? Why?