Algebra I
Unit 6: Describing Data

References

HMH Georgia Coordinate Algebra Text:
Unit 4: Modules 14 & 15
HMH Georgia Analytic Geometry Text:
Unit 5: Module 15.3
(quadratic regression)

Check with you teacher for online and print access:

Online website:
my.hrw.com

Web Resources

- GA Virtual: Interpreting and Representing Two Variable Data
http://cms.gavirtualschool.org/Shared/Math/GSEAlgI16/GSEAlgI_IntandRepTwoVarData_Shared/index.html
- Khan Academy
https://www.khanacademy.org/math/probability/regression
- Correlation Coefficient
http://mathbits.com/MathBits/TISection/Statistics2/correlation.htm
- Two-way Frequency Tables
https://mathbitsnotebook.com/Algebra1/StatisticsReg/ST2TwoWayTable.html
- Shapes of Distributions
http://www.mathbitsnotebook.com/Algebra1/StatisticsData/STShapes.html

Dear Parents

Below you will find a list of concepts that your child will use and understand while completing Unit 6: Describing Data. Also included are references, vocabulary and examples that will help you assist your child at home.

Concepts Students will Use and Understand

- Know how to compute the mean, median, interquartile range, and mean absolute deviation by hand in simple cases and using technology with larger data sets.
- Find the lower extreme (minimum), upper extreme (maximum), and quartiles.
- Use and interpret shape, center, and spread of data.
- Create a graphical representation of a data set.
- Summarize data in two-way frequency table.
- Represent data in a scatter plot and describe how the variables are related.
- Interpret the slope & y-intercept of a line from any representation.
- Find linear, quadratic, and exponential regressions.
- Compute and interpret the correlation coefficient.
- Understand the meaning of correlation and causation.

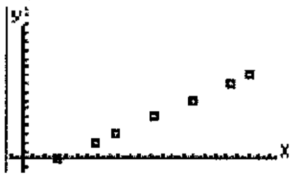
Vocabulary

- **Bivariate data.** Pairs of linked numerical observations. Example: a list of heights and weights for each player on a football team.
- **Conditional Frequencies.** The relative frequencies in the body of a two-way frequency table.
- **Correlation Coefficient.** A measure of the strength of the linear relationship between two variables that is defined in terms of the (sample) covariance of the variables divided by their (sample) standard deviations.
- **Joint Frequencies.** Entries in the body of a two-way frequency table.
- **Marginal Frequencies.** Entries in the "Total" row and "Total" column of a two-way frequency table.
- **Mean Absolute Deviation.** A measure of variation in numerical data by adding the distance between each data point and the mean, then dividing by the number of values
- **Shape**. The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity.
- **Symmetry**. A symmetric distribution can be divided at the center so that each half is a mirror image of the other.
- **Number of Peaks**. Distributions can have few or many peaks. Distributions with one clear peak are called unimodal and distributions with two clear peaks are called bimodal. Unimodal distributions are sometimes called bell-shaped.
- **Direction of Skew**. Some distributions have many more observations on one side of graph than the other. Distributions with a tail on the right toward the higher values are said to be skewed right; and distributions with a tail on the left toward the lower values are said to be skewed left.
- **Uniformity**- When observations in a set of data are equally spread across the range of the distribution, the distribution is called uniform distribution. A uniform distribution has no clear peaks.
- **Trend.** A change (positive, negative or constant) in data values over time.

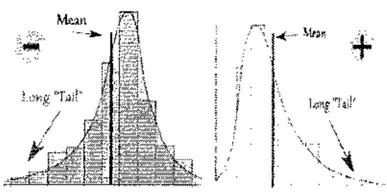
Practice Problems

Answers

1.) $y = 0.556x - 17.778$



2.) When you look at the shape of the data, if the "long tail" is on the left=skewed left, if it is on the right=skewed right, and if it is evenly distributed it is symmetric.



3.) The correlation coefficient is approximately .999. This means the line of best fit is extremely accurate because the coefficient is so close to 1.

4.) table in yellow at bottom is the key

a. 36%, 19%, 45%, 54%, 46%

b. 25%, 9%, 20%(top), 11%, 10%, 25%(bottom)

c. 54%

d. 36%

e. 69.4%

1.) Find the linear regression of the following data:

Fahrenheit degrees (°F)	Celsius degrees (°C)
32	0
68	20
86	30
122	50
158	70
194	90
212	100

2.) Explain when data is skewed left, right, or symmetric.

3.) Using technology, determine the correlation coefficient. Interpret its meaning.
(0,20) (1,40) (2,75) (3,150) (4, 297) (5,510)

4.) Construct a frequency table from the following information:
A survey of 200 9th and 10th graders was given to determine what their favorite subject was. 72 said Math (50 which were freshmen), 38 said Social Studies (20 which were sophomores), and 40 freshmen and 50 sophomores said PE was their favorite.

Based on your tables above, answer the following questions:

a. What are the marginal relative frequencies?
b. What are the joint relative frequencies?
c. What is the probability that a student surveyed is a freshman?
d. What is the probability that a student surveyed likes Math?
e. If a student likes Math, what is the probability that they are a freshman?

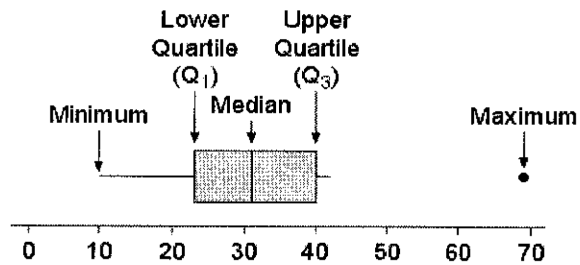
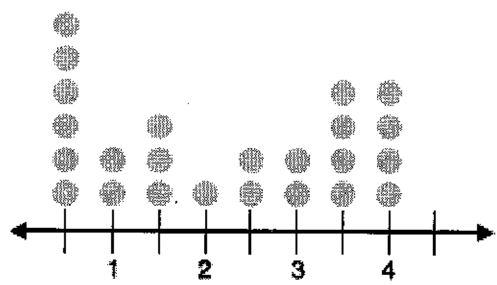
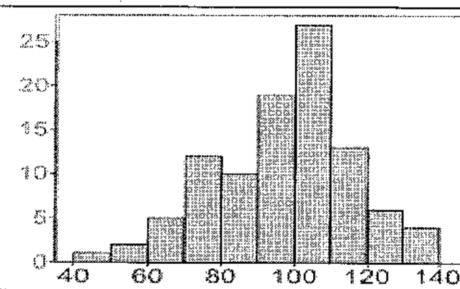
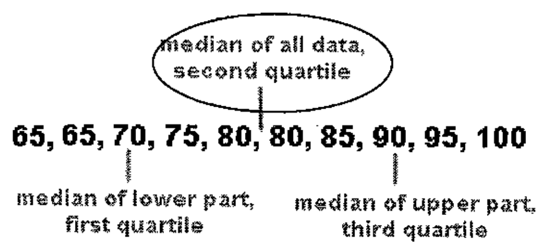
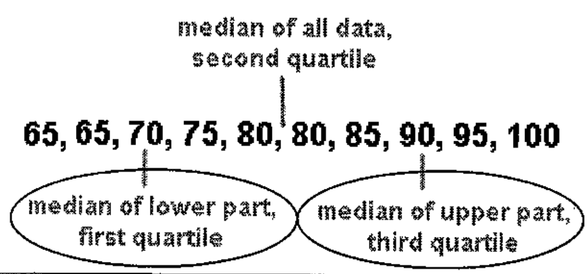
	Math	SS	PE	Total
9th	50	18	40	108
10th	22	20	50	92
Total	72	38	90	200

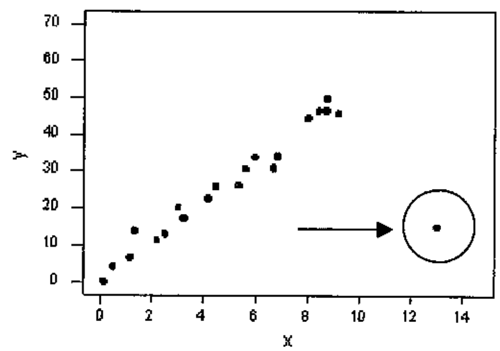
Algebra I Unit 6: Describing Data

Student Friendly Standards

Standard	Mastery Level	Statements
S.ID.1		☐ I can represent data with plots on the real number line (dot plots, histograms, and box plots). ☐ I can choose appropriate graphs to be consistent with numerical data: dot plots, histograms, and box plots.
S.ID.2		☐ I can use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, standard deviation) of two or more different data sets.
S.ID.3		☐ I can interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers). ☐ I can examine graphical representations to determine if data are symmetric, skewed left, or skewed right and how the shape of the data affects descriptive statistics.
S.ID.5		☐ I can summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). ☐ I can recognize possible associations and trends in the data.
S.ID.6		☐ I can represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
S.ID.6a		☐ I can decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function to best fit. ☐ I can use this function to solve problems in context. ☐ I can emphasize linear, quadratic, and exponential models.
S.ID.6c		☐ I can fit a linear function for a scatter plot that suggests a linear association.
S.ID.7		☐ I can interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.
S.ID.8		☐ I can compute (using technology) and interpret the correlation coefficient "r" of a linear fit. ☐ After calculating the line of best fit using technology, I can describe how strong the goodness of fit of the regression is, using "r."
S.ID.9		☐ I can distinguish between correlation and causation.

Day 1 and 2 Introduction to Statistics - Vocabulary

Term	Describe	Example
Box Plot		
Dot Plot		
Histogram		
Median		
First and Third Quartiles		

Interquartile Range		**Subtract** **Third Quartile (Q_3) – First Quartile (Q_1) = IQR**
Outlier		
Mean		**5 + 4 + 2 + 6 + 3 = 20** $\frac{20}{5} = 4$ **The Mean is <u>4</u>.**
Mean Absolute Deviation (MAD)		Steps: 1. Find the Mean 2. Calculate the absolute value of the difference between each data value and the mean 3. Determine the average of the differences in step 2. This average is the mean absolute deviation
Measures of Center		Find the Mean and Median for the following data. ***<u>Hint:</u>*** (Must order the numbers first before finding the Median) 2 1 5 4 3 **<u>Mean</u>**: $\frac{15}{5} = 3$ **<u>Median</u>** = 3
Measures of spread		***<u>Examples of Measures of Spread:</u>*** 1. Range 2. Interquartile Range (IQR) 3. Mean Absolute Deviation -MAD

Graphic Organizer: Measures of Center and Spread

Mean

Median

Measures of Center

Measures of Spread

Mean Absolute Deviation

Interquartile Range

Name______________________________ Date__________ Class________

Guided Notes – SP.4

The _______ , __________ and ________ of a data set are used to measure where the center of a set of data lies. Other measures indicate how spread out, or how ____________, data are. These measures include ________, the interquartile range (_____), and the mean absolute deviation (_____).

Measures of Center

________ is the average or sum of all data points divided by the number of points.

Example: Look at the data set: 3, 7, 4, 9, 6

1) Find the sum of all the numbers __________
2) Count how many data points are in the set _________
3) Divide the sum by the number of data points ________________

___________ is the middle value when the data points are in increasing order. This is also known as the 2nd quartile (Q_2).

Example 1: Look at the data set: 3, 7, 4, 9, 6

1) Put the data in order from least to greatest ________________
2) Find the middle number ____________________

Example 2: Look at the data set: 3, 7, 9, 6

1) Put the data in order from least to greatest ________________
2) If there is no middle number, find the 2 middle numbers __________
3) Add the numbers together and divide by 2 ____________________

__________ is the most often occurring number(s), if all the numbers are listed the same amount of times, there is no ___________.

Measures of Variability

___________ of a set of data is the difference between the largest and the smallest number.

Example: Look at the data set: 3, 7, 4, 9, 6, 2, 5

1) Put the data in order from least to greatest ________________
2) Subtract the smallest number from the largest ________________

______________________ is the middle of the lower half of the data set.

Example: Look at the data set: 3, 7, 4, 9, 6, 2, 5

1) Put the data in order from least to greatest ________________
2) Circle the median
3) Find the median of the lower half of the data and circle ____________

______________________ is the middle of the upper half of the data set.

Example: Look at the data set: 3, 7, 4, 9, 6, 2, 5

1) Put the data in order from least to greatest ________________
2) Circle the median
3) Find the median of the upper half of the data and circle ____________

________________________ is the value of the 1st quartile (Q_1) subtracted from the value of the 3rd quartile (Q_3) in a data set.

Example: Look at the data set: 3, 9, 2, 6, 4, 7, 3, 8

1) Put the data in order from least to greatest ________________
2) Find the median (Q_2) ________________
3) Find the 1st quartile (Q_1) ________________
4) Find the 3rd quartile (Q_3) ________________
5) Subtract the 1st quartile from the 3rd quartile ________________

________________________ is the average of how much the data points in a set deviate or vary from the mean. Since distance is always positive, you must take the absolute value of each deviation.

Example: Look at the data set: 3, 9, 2, 6, 4, 7, 3, 8

1) Put the data in order from least to greatest ________________
2) Find the mean ________________
3) Find the absolute deviation of each data point from the mean. Use the table below to organize your work.

Data Point	Deviation from Mean	Absolute Deviation from Mean
2		
3		
3		
4		
6		
7		
8		
9		

4) Calculate the mean of the absolute deviations ________________
5) The mean absolute deviation (MAD) is ________

Using the data set below, find the mean, median, mode, range, 1st quartile, 3rd quartile, interquartile range, and mean absolute deviation.

3, 5, 7, 7, 8, 12, 13, 14, 18, 18, 21

Mean ________________ Median ________________

Mode ________________ Range ________________

Q_1 ________________ Q_3 ________________

IQR ________________ MAD ________________

Measures of Spread – Range, IQR, and Mean Absolute Deviation

Mean Absolute Deviation of a numerical data set is the average positive deviations of the data from the mean.

$$\textit{Mean Absolute Deviation} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{n}$$

A measure of distribution is a measure of how spread out data is, or how the data is distributed from its smallest values to its largest values. Suppose, for instance, that Joe has test scores of 60, 68, 69, 78, 90, 95, and 100. Sammy scores 78, 78, 79, 79, 82, 82, and 82.

8. Calculate Joe's mean test score. Then calculate Sam's mean test score. What do you notice about Joe's scores compared to Sammy's?

Measuring the mean will not tell you much about the characteristics of the test takers. A measure of distribution, or spread, will help you see that Sam consistently scores near 80, while Joe's scores are spread out, or distributed, over a much larger range.

9. To examine the distribution of test scores, find the mean absolute deviation. Follow the steps below to find the mean absolute deviation of Sam's test scores (Joe's example is given).

Steps	Joe	Sam
a) Calculate the mean, symbolically, $\bar{x}$, of the data.	Mean: $\bar{x} = \frac{60+68+69+78+90+95+100}{7}$, $\bar{x} = 80$	
b) Find the deviation, or distance from the mean, for each piece of data.	Deviation: $60-80=-20$ $68-80=-12$ $69-80=-11$ $78-80=-2$ $90-80=10$ $95-80=15$ $100-80=20$	
c) Find the absolute value of the mean deviations.	20 12 11 2 10 15 20	
d) Find the average of the positive deviations found in part c.	MAD: $\frac{20+12+11+2+10+15+20}{7} = 12.86$	

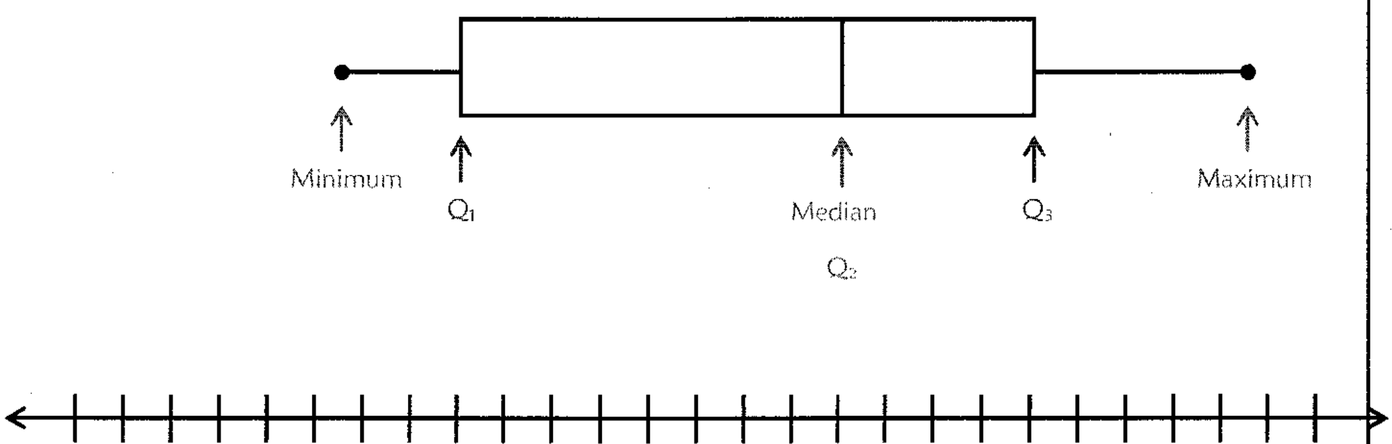
10. Why do you think the MAD of Joe's test scores is higher than the MAD of Sammy's test scores?

KEY CONCEPTS AND VOCABULARY

A ______________________________ is a bar graph that is used to display the frequency of data divided into equal intervals. The heights of the bars indicate the frequency of data values. The bars must be of equal width and should touch but not overlap.

A ____ can be used to show how the values in a data set are distributed. You need 5 values to make a box plot: the minimum, first quartile, median, third quartile, and maximum.

BOX-AND-WHISKER PLOT



OUTLIER

An outlier is value in a data set that is much greater or much less than most of the other values in the data set.

A data value x is an outlier if:

$$x < Q_1 - 1.5(IQR) \text{ or } x > Q_3 + 1.5(IQR)$$

EXAMPLE 5: COMPARING BOX-AND-WHISKER PLOTS

The bowling averages for two competitive bowling teams are given in the table.

Bowling Team 1	202	145	172	182	167	145	150	132	128
Bowling Team 2	192	211	185	173	178	166	158	171	120

a) Use the data to make a box plot for each team on the same number line.

b) Compare the distributions of the bowling averages for each team.

c) Which team is better? Explain.

RATE YOUR UNDERSTANDING *(Using the learning scale from the beginning of the lesson)*

Circle one: 4 3 2 1

Chapters 8 & 10 Name ______________________ Period ________
Probability & Statistics

It isn't always easy to read and interpret the data listed in a table. Sometimes, a better way to organize data is to create a graph. In this chapter, we will look at different graphs and how the data is displayed.

DOT PLOTS

A **dot plot** distributes discrete data using a number line. **Discrete data** has a finite value that can be measured or counted, like the number of calories in a candy bar.

Dot plots are best used to organize and display the number of occurrences for a small amount of data.

Cereal Name	Sugar in one Serving (grams)
Cocoa Rounds	13
Flakes of Corn	4
Frosty Flakes	11
Grape Nuggets	7
Golden Nuggets	10
Honey Nut Squares	10
Raisin Branola	7
Healthy Living Flakes	7
Wheatleys	8
Healthy Living Crunch	6
Multi-Grain Squares	7
All Branola	5
Munch Crunch	12
Branola Flakes	5
Complete Flakes	4
Corn Crisps	3
Rice Crisps	4
Shredded Wheatleys	1
Puffs	22
Fruit Circles	11

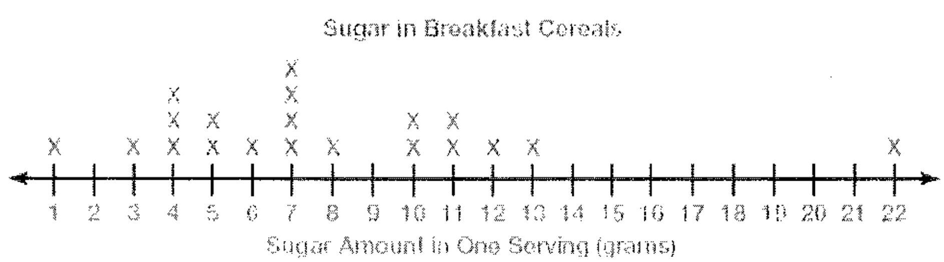
Lowest data value = 1

Highest data value = 22

So, the number line must include the numbers between 1 and 22.

Construct a dot plot:

1) Draw a number line, marking intervals that include all the data values listed in the table.
2) Place an "x" above the number that represents each data value. The "x"s represent the number of times that data value is listed in the table.
3) Title the dot plot and identify the data values and their unit of measure.



12

EXAMPLES

EXAMPLE 1: USING DOT PLOTS TO REPRESENT DATA

Create a dot plot for each the data set.

a) The results of a survey on the number of television per household.

2	2	3
4	1	3
2	3	4
5	3	1
2	3	4

b) The number of students in each biology class in a high school.

25	26	23
27	25	22
24	23	26
23	24	21

EXAMPLE 2: IDENTIFYING OUTLIERS AND INVESTIGATING THE EFFECT OF AN OUTLIER TO A DATA SET

Use the following data to complete the following problems.

12	17	15	10	16	10
17	19	10	18	14	16

a) Create a dot plot for the data set using an appropriate scale for the number line.

b) If 43 was added to the data, would it be an outlier?

c) If 1 was added to the data, would it be an outlier?

EXAMPLE 3: COMPARING DATA SETS USING MEASURES OF CENTRAL TENDENCY

The tables describe the ages of employees from two randomly chosen companies.

COMPANY 1	35	37	40	65	26	55	57	38	42
COMPANY 2	23	22	28	35	31	47	50	19	35

a) Calculate the mean, median, interquartile range (IQR), and standard deviation for each data set.

	MEAN	MEDIAN	IQR	STANDARD DEVIATION
COMPANY 1				
COMPANY 2				

b) Compare the data sets.

EXAMPLE 4: COMPARING DATA SETS USING DOT PLOTS

The table describes the points each player makes for two basketball teams in a game.

TEAM A	12	10	10	8	12	4	14	14	16	14	2
TEAM B	4	4	6	34	6	2	2	4	6	20	4

a) For each data set, make a dot plot and determine the type of distribution.

b) Explain what the distribution means for each data set.

c) Are there any outliers for either team? What does this mean for the data?

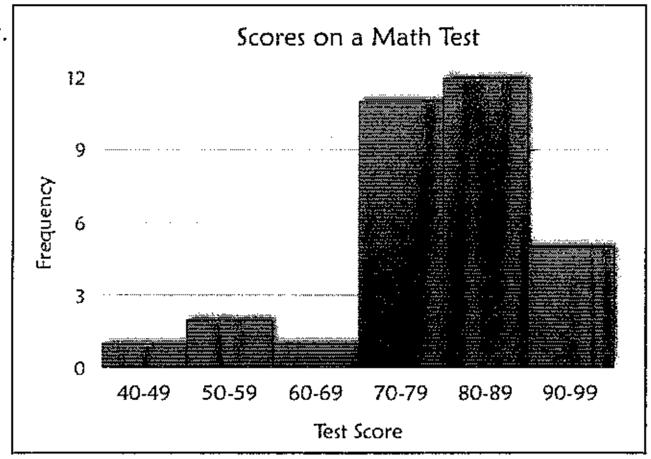
RATE YOUR UNDERSTANDING *(Using the learning scale from the beginning of the lesson)*

Circle one: 4 3 2 1

EXAMPLES

EXAMPLE 1: UNDERSTANDING HISTOGRAMS

Use the histogram to answer the following questions.



a) How many test takers received a score between 50-59?

b) What does the horizontal axis represent?

c) How many test takers received a passing score on the test (60 or higher)?

d) How many test takers failed (less than 60)?

e) How many people took the test?

EXAMPLE 2: CREATING HISTOGRAMS

Create a histogram for the given data.

a) Listed are the ages are the United States presidents on their inauguration day.

57	61	57	57	58	57	61	54	68	51	49	64	50
48	65	52	56	46	54	49	51	47	55	55	54	42
51	56	55	51	54	51	60	62	43	55	56	61	52
69	64	46	54	47								

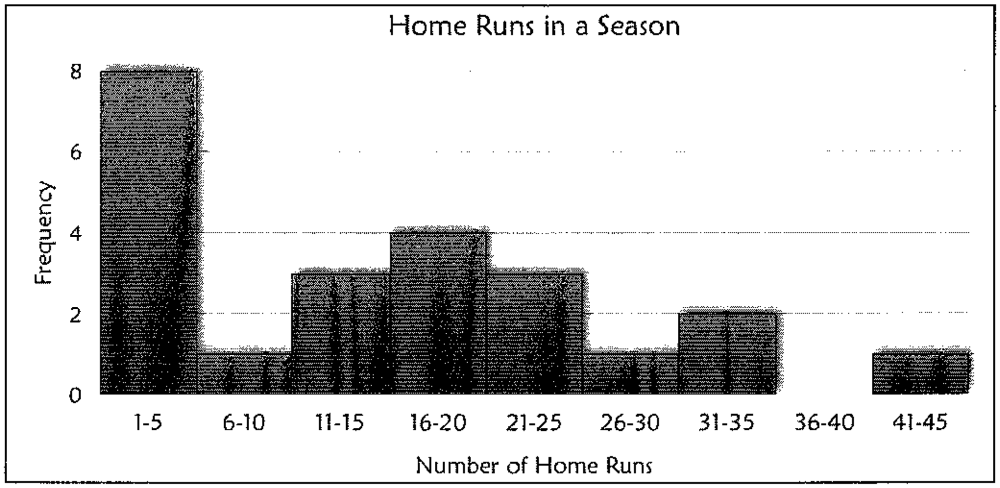
b) Listed are the average points per game for players of a basketball team.

27.1	9.6	7.9	6.6	4.9	4.5	4	4.1
3.8	3.3	2.9	9.8	16.2	6.4	19	

EXAMPLE 3: ESTIMATING DATA USING HISTOGRAMS

The histogram shows the number of home runs each player on a baseball team hit in a particular season.

a) Estimate the mean of the data set displayed in the histogram.



b) What does the mean represent in context of the problem?

EXAMPLE 4: CREATING BOX-AND-WHISKER PLOTS

Use the data to make a box plot.

a) Listed are the scores from a professional golf tournament.

68	76	71	69	75	76	74	77
78	73	70	76	74	76	75	71

b) Listed is the number of cars that park in a parking garage at a mall for twelve days.

35	57	103	138	110	45
57	49	62	98	145	106

16

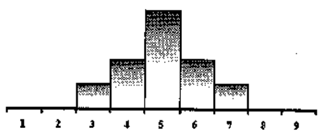
Notes: Comparing Distributions

When you compare two or more data sets, focus on four features:

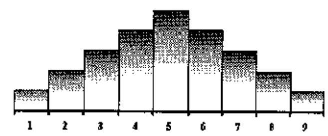
- **Center.** Graphically, the center of a distribution is the point where about half of the observations are on either side.
- **Spread**. The spread of a distribution refers to the variability of the data. If the observations cover a wide range, the spread is larger. If the observations are clustered around a single value, the spread is smaller.
- **Shape**. The shape of a distribution is described by symmetry, skewness, number of peaks, etc.
- **Unusual features**. Unusual features refer to gaps (areas of the distribution where there are no observations) and outliers.

SPREAD

The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set.



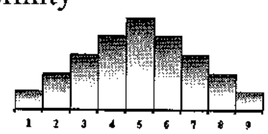
Less Spread



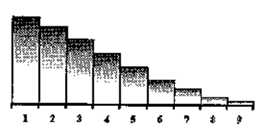
More Spread

SHAPE

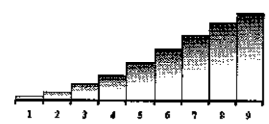
The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity



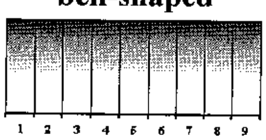
symmetric, unimodal, bell-shaped



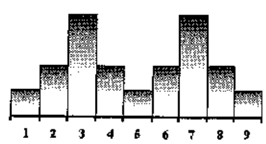
skewed right



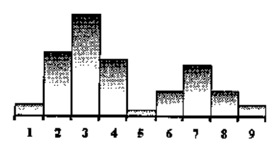
skewed left



uniform



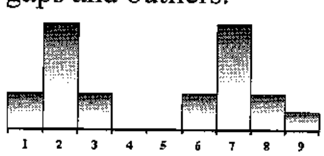
symmetric, bimodal



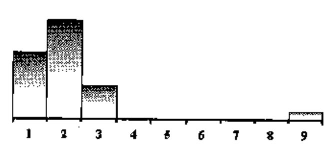
non-symmetric, bimodal

UNUSUAL FEATURES

Sometimes, statisticians refer to unusual features in a set of data. The two most common unusual features are gaps and outliers.



gap



outlier

RELATIVE FREQUENCY

MACC.912.S-ID.B.5: Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

RATING	LEARNING SCALE
4	I am able to • identify and analyze joint, marginal, and conditional relative frequency in more challenging problems that I have never previously attempted
TARGET 3	I am able to • identify and analyze joint, marginal, and conditional relative frequency
2	I am able to • identify and analyze joint, marginal, and conditional relative frequency with help
1	I am able to • know the definition of relative frequency

WARM UP

Convert each fraction to a percent.

a) $\frac{16}{25}$ b) $\frac{31}{45}$ c) $\frac{13}{20}$ d) $\frac{37}{80}$

KEY CONCEPTS AND VOCABULARY

The ______________________________ of a category is the frequency of the category divided by the total of all frequencies.

Types of Relative Frequencies

- Joint Relative Frequency is found by dividing a frequency that is not in the total row or the total column by the grand total.
- Marginal Relative Frequency is found by dividing a row total or a column total by the grand total.
- Conditional Relative Frequency is found by dividing a frequency that is not in the Total row or the Total column by the total for that row or column.

EXAMPLES

EXAMPLE 1: COMPLETING A RELATIVE FREQUENCY TABLE

Students were surveyed regarding their favorite school subject.

	FAVORITE SUBJECT – FREQUENCY TABLE				
FAVORITE SUBJECT	MATH	SCIENCE	ENGLISH	HISTORY	TOTAL
FREQUENCY	7	6	8	4	25

a) Complete the relative frequency table for this data using percents rounded to the nearest tenth.

	FAVORITE SUBJECT – RELATIVE FREQUENCY TABLE				
FAVORITE SUBJECT	MATH	SCIENCE	ENGLISH	HISTORY	TOTAL
RELATIVE FREQUENCY					

b) What is the relative frequency of having history as a favorite subject?

EXAMPLE 2: COMPLETING A TWO-WAY RELATIVE FREQUENCY TABLE

Students were surveyed regarding their favorite out of blue, green, red, and orange. Their gender was also recorded.

	FAVORITE COLOR – TWO-WAY FREQUENCY TABLE				
GENDER	BLUE	RED	GREEN	ORANGE	TOTAL
MALE	12	16	9	3	40
FEMALE	17	10	8	5	40
TOTAL	29	26	17	8	80

a) Complete a two-way relative frequency table from the data in a two-way frequency table.

	FAVORITE COLOR– RELATIVE FREQUENCY TABLE				
GENDER	BLUE	RED	GREEN	ORANGE	TOTAL
MALE					
FEMALE					
TOTAL					

b) Identify the joint relative frequencies.

c) Identify the marginal relative frequencies.

EXAMPLE 3: CALCULATING AND INTERPRETING CONDITIONAL RELATIVE FREQUENCY

Use the joint relative frequencies from example 2 to calculate the associated conditional relative frequencies and describe what each one means.

a) Find the conditional relative frequency that a student is a boy, given that the student prefers red.

b) Find the conditional relative frequency that a student prefers blue, given that the student is a girl.

c) Find the conditional relative frequency that a student is a girl, given that the student prefers green.

EXAMPLE 4: ANALYZING A RELATIVE FREQUENCY TABLE

Seventy-five students were surveyed about working a summer job. The results are shown in the table.

	WORKING A SUMMER JOB TWO-WAY FREQUENCY TABLE		
GENDER	YES	NO	TOTAL
MALE	27	13	40
FEMALE	21	14	35
TOTAL	48	27	75

a) Complete a two-way relative frequency table from the data in a two-way frequency table.

	WORKING A SUMMER JOB RELATIVE FREQUENCY TABLE		
GENDER	YES	NO	TOTAL
MALE			
FEMALE			
TOTAL			

b) What percent of the students surveyed work a summer job? Identify the type of relative frequency.

c) What percent of the students surveyed are female and work a summer job? Identify the type of relative frequency.

d) What percent of the males surveyed do not work a summer job? Identify the type of relative frequency.

e) What percent of the males surveyed work a summer job? Identify the type of relative frequency.

f) Is there an association between the gender of a student and whether they work a summer job? Explain.

***RATE YOUR UNDERSTANDING** (Using the learning scale from the beginning of the lesson)*

Circle one: 4 3 2 1

Name ______________________ Date ____________ Class____________

LESSON 14-3

Practice B

Two-Way Tables

1. The table shows the results of a customer satisfaction survey of 100 randomly selected shoppers at the mall who were asked if they would shop at an earlier time if the mall opened earlier. Make a table of joint and marginal relative frequencies.

	Ages 10–20	Ages 21–45	Ages 46–65	65 and Older
Yes	13	2	8	24
No	25	10	15	3

	Ages 10–20	Ages 21–45	Ages 46–65	65 and Older	Total
Yes					
No					
Total					

2. Jerrod collected data on 100 randomly selected students, and summarized the results in a table.

Owns an MP3 Player

		Yes	No
Owns a Smart phone	**Yes**	28	12
	No	34	26

a. Make a table of the joint relative frequencies and marginal relative frequencies. Round to the nearest hundredth where appropriate.

Owns an MP3 player

		Yes	No	Total
Owns a Smart Phone	**Yes**			
	No			
	Total			

b. If you are given that a student owns an MP3 player, what is the probability that the student also owns a smart phone? Round your answer to the nearest hundredth.

____

c. If you are given that a student owns a smart phone, what is the probability that the student also owns an MP3 player? Round your answer to the nearest hundredth.

____

21

Name ______________________ Date ______________ Class______________

LESSON 14-3

Problem Solving

Two-Way Tables

1. The table shows the number of students who would drive to school if the school provided parking spaces. Make a table of joint relative frequencies and marginal relative frequencies.

	Lowerclassmates	Upperclassmates
Always	32	122
Sometimes	58	44
Never	24	120

	Lowerclassmates	Upperclassmates	Total
Always			
Sometimes			
Never			
Total			

2. Gerry collected data and made a table of marginal relative frequencies on the number of students who participate in chorus and the number who participate in band.

		Chorus		
		Yes	**No**	**Total**
Band	**Yes**	0.38	0.29	0.67
	No	0.09	0.24	0.33
	Total	0.47	0.53	1.0

a. If you are given that a student is in chorus, what is the probability that the student also is in band? Round your answer to the nearest hundredth.

____

b. If you are given that a student is not in band, what is the probability that the student is in chorus? Round your answer to the nearest hundredth.

____

Select the best answer.

3. What is the probability if a student is not in chorus, then that student is in band?

A 0.29 B 0.38

C 0.43 D 0.55

4. What is the probability that if a student is not in band, then that student is not in chorus?

F 0.09 G 0.33

H 0.44 J 0.73

22

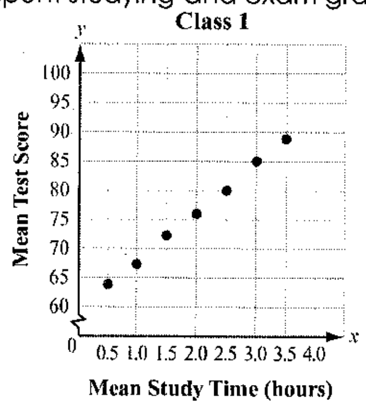
Name: ______________________________ Date: ________ Period__________

Correlation

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MCC9-12.S.ID.9 Distinguish between correlation and causation.

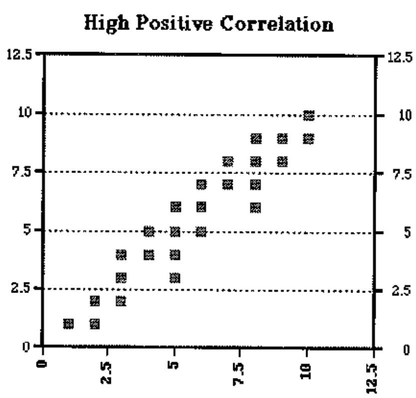
A **scatter plot** is often used to present bivariate **quantitative** data. Each variable is represented on an axis and the axes are labeled accordingly.

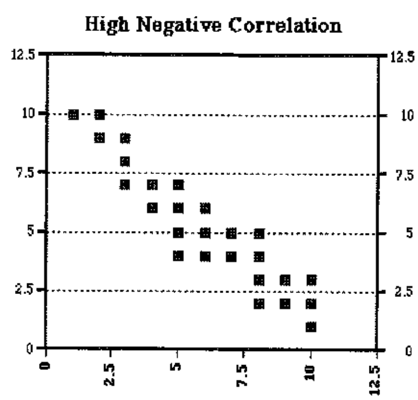
A scatter plot displays data as points on a grid using the associated numbers as coordinates or ordered pairs (x, y). The way the points are arranged by themselves in a scatter plot may or may not suggest a relationship between the two variables. For instance, by reading the graph below, do you think there is a relationship between the hours spent studying and exam grades?

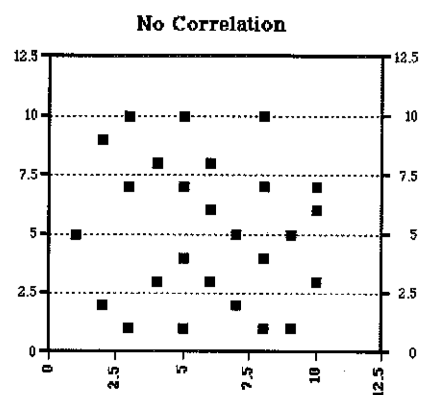


If y tends to increase as x increases, then the data have **positive** correlation.

If y tends to decrease as x increases, then the data have **negative** correlation.







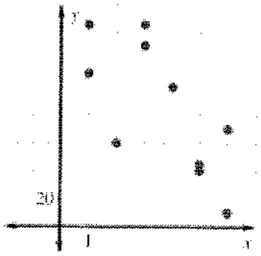
A correlation coefficient, denoted by r, is a number from -1 to 1 that measures how well a line fits a set of data pairs (x, y). If r is near 1, the points lie close to a line with a positive slope. If r is near -1, the points lie close to a line with a negative slope. If r is near 0, the points to not lie close to any line.

Give an example of negative correlation: ______________________________

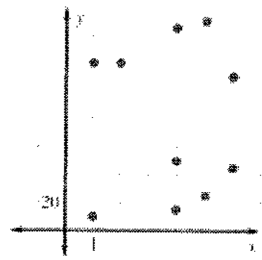
<u>Practice Problems</u>:

For each scatter plot, tell whether the data have a positive correlation, a negative correlation, or no correlation. Then, tell whether the correlation is closest to -1, -0.5, 0, 0.5, or 1.

1.



2.



3. Positive, negative, or no correlation?

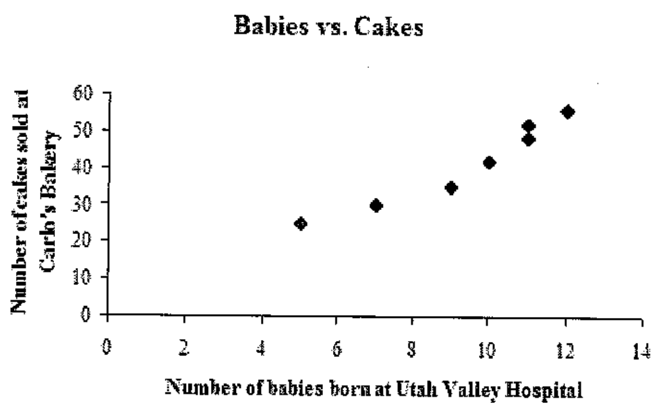
a. Amount of exercise and percent of body fat ___________
b. A person's age and the number of medical conditions they have ___________
c. Temperature and number of ice cream cones sold ___________
d. The number of students at Hillgrove and the number of dogs in Atlanta ___________
e. Age of a tadpole and the length of its tail ___________

Correlation vs. Causation

When a scatter plot shows a correlation between two variables, even if it's a strong one, there is <u>*not*</u> *necessarily a cause-and-effect relationship*. Both variables could be related to some third variable that actually causes the apparent correlation. Also, an apparent correlation simply could be the result of chance.

<u>Example 1</u>: During the month of June the number of new babies born at the Utah Valley Hospital was recorded for a week. Over the same time period, the number of cakes sold at Carlo's Bakery in Hoboken, New Jersey was also recorded. What can be said about the correlation? Is there causation? Why or why not?

Number of babies born	Number of cakes sold
5	25
7	30
9	35
10	42
11	48
11	52
12	56



<u>Example 2</u>: Mr. Jones gave a math test to all the students in his school. He made the startling discovery that the taller students did better than the short ones. His Causation Statement: *As your height increases, so does your math ability.*
What can be said about the correlation? Is there causation? Why or why not?

<u>Example 3</u>: In this present economy families are trying to find ways to save money Families might be thinking about not eating out to spend less money. Causation Statement: *The more you eat out, the more money you spend at restaurants.*
What can be said about the correlation? Is there causation? Why or why not?

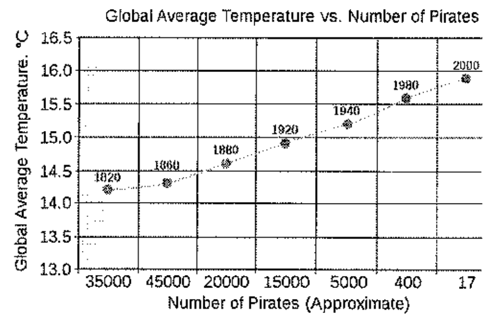
Name: ______________________________ Date: ________Period__________

Correlation and Causation Homework

1. From the information given,
   a. Determine if the correlation is positive, negative or none.

   b. Estimate the correlation coefficient.

   c. Is there causation? Why or why not?



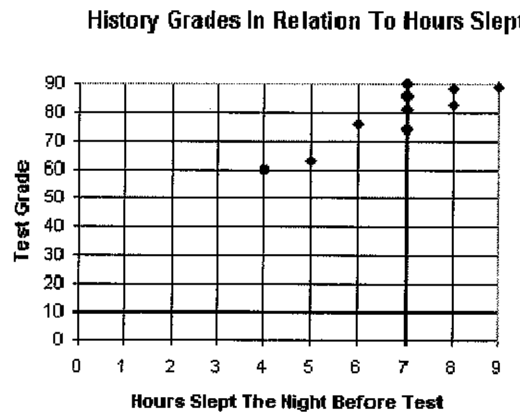
---

2. A history teacher asked her students how many hours of sleep they had the night before a test. The data above shows the number of hours the student slept and their score on the exam. The graph is a scatter plot from the given data.

   a. Determine if the correlation is positive, negative, or none.

   b. Estimate the correlation coefficient.

   c. Is there causation? Would this information affect your behavior the night before a test?

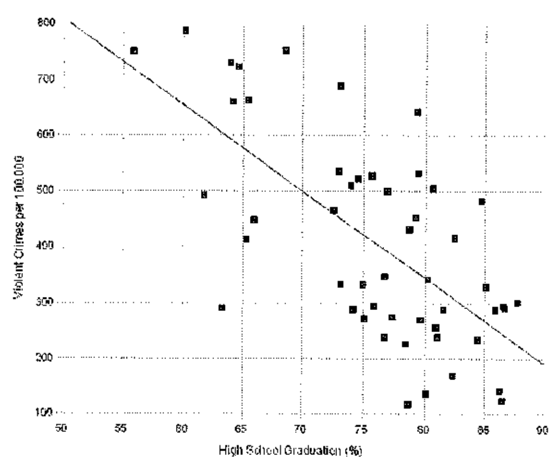


---

3. The following chart shows violent crime rates compared to high school graduation for all fifty states.
   a. Determine if the correlation is positive, negative, or none.

   b. Estimate the correlation coefficient.

   c. Is this an illustration of cause and effect, or are these two variables simply correlated?



Adapted from: Mathematics Vision Project

For the given situations below,

a. Is the association positive, negative or none?
b. Is the causation statement is true or false?

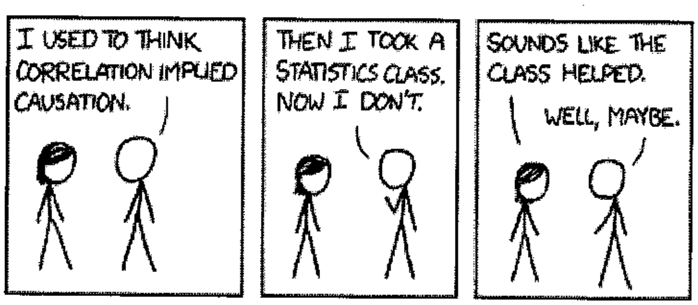
4. When you are on a diet, the less calories you eat daily vs. the more weight you lose. Causation statement: *Therefore, eating less calories makes you lose weight.*

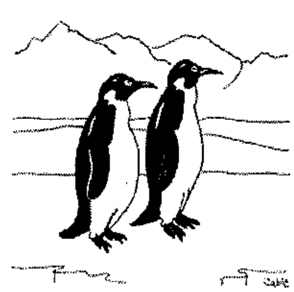
5. The more ice cream consumed on a beach vs. the increased number of people who go in the water. Causation statement: *Therefore, eating more ice cream on the beach makes people go in the water.*

6. The more people in a family vs. the increased number of cars the family owns. Causation Statement: Therefore, the more people there are in a family determines how many cars a family owns.

7. The average speed cars travel from Philadelphia to New York on the turnpike vs. the average amount of times it takes. Causation Statement: *Therefore, the speed cars travel from Philadelphia to New York determines the time it takes to go between them.*

8. How much you pay for a house vs. how much you pay for a car. Causation statement: *Therefore the more you pay for a house makes you spend more for a car.*





"Do you think all these film crews brought on global warming or did global warming bring on all these film crews?"

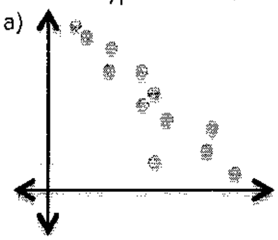
Adapted from: Mathematics Vision Project

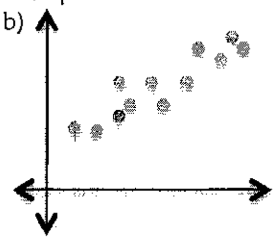
~~39~~ 26

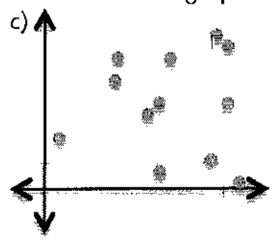
EXAMPLES

EXAMPLE 1: IDENTIFYING CORRELATION AND ESTIMATING THE CORRELATION COEFFICIENT

Describe the type of correlation the scatterplot shows. Estimate the value of r for each graph.

a) 

b) 

c) 

EXAMPLE 2: WRITING AN EQUATION OF A LINE OF BEST FIT

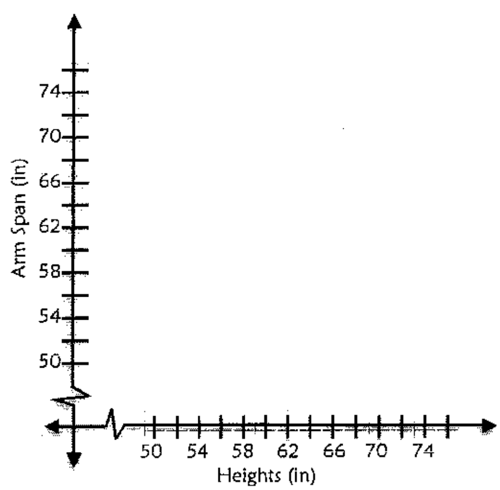
Use the data to make a scatter plot for the data below.

Heights and Arm Spans									
Height (in)	63	70	60	62	64	65	72	59	61
Arm Span (in)	62	67	60	61	63	65	70	59	60

a) Draw a line of best fit.

b) Estimate the correlation coefficient.

c) Find the equation for the line of best fit.



d) Estimate the arm span of a person who is 67 inches tall.

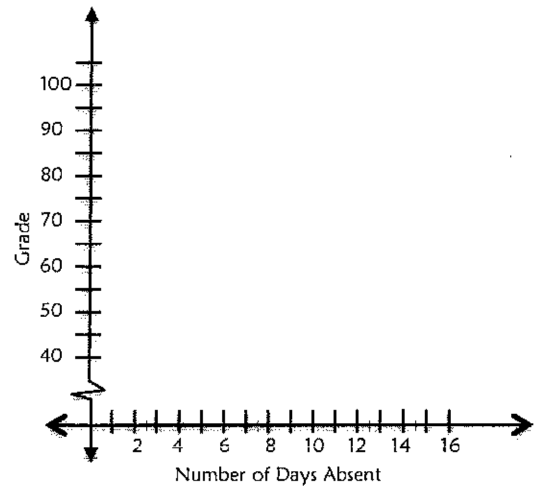
e) Estimate the height of a person who has an arm span of 48 inches

EXAMPLE 3: INTERPRETING A LINE OF BEST FIT

Use the data to make a scatter plot for the data below.

Grades and Number of Absences											
Number of Absences (days)	2	1	12	8	0	4	5	7	15	2	3
Grade	88	90	55	61	96	80	70	75	52	93	83

a) Draw a line of best fit.



b) Estimate the correlation coefficient.

c) Find the equation for the line of best fit.

d) Estimate the grade for a student who has missed 10 days of school.

e) Using the line of best fit from part c, what is the x-intercept? What does it mean in context of the problem.

f) Using the line of best fit from part c, what is the slope? What does it mean in context of the problem.

EXAMPLE 4: DISTINGUISHING BETWEEN CORRELATION AND CAUSATION

Read each description. Identify the variables in each situation and determine whether it describes a positive or negative correlation. Explain whether the correlation is a result of causation.

a) The average air temperature and how much ice cream an ice cream shop sells on any given day.

b) The cost of a family vacation and the size of their house.

***RATE YOUR UNDERSTANDING** (Using the learning scale from the beginning of the lesson)*

Circle one: 4 3 2 1

Name: ______________________________ Date: ___________Period_______

Scatter Plots and Line of Best Fit

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use the given functions or choose a function suggested by the context. Emphasize linear and exponential models.
MCC9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

The **best fitting line or curve** is the line that lies as close as possible to all the data points.

Regression is a method used to find the equation of the best fitting line or curve.

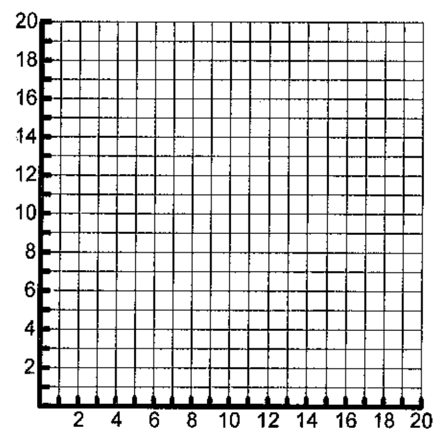
Extrapolation – the use of the regression curve to make predictions outside the domain of values of the independent variable.

Interpolation – Interpolation is used to make predictions within the domain of values of the independent variable.

Line of Best Fit by Hand:

1) The environment club is interested in the relationship between the number of canned beverages sold in the cafeteria and the number of cans that are recycled. The data they collected are listed in this chart.

Beverage Can Recycling								
Number of Canned Beverages Sold	18	15	19	8	10	13	9	14
Number of Cans Recycled	8	6	10	6	3	7	5	4



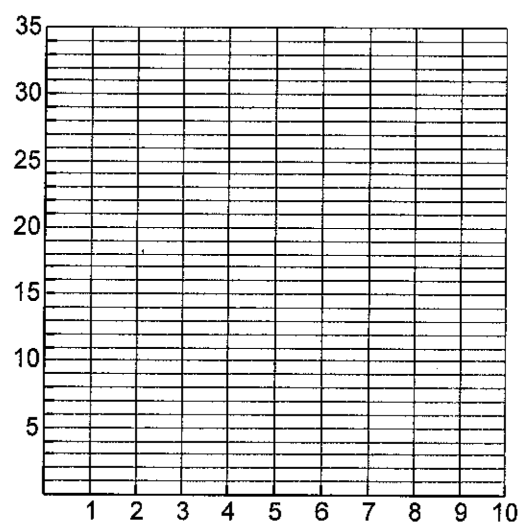
a) Plot the points to make a scatter plot.

b) Use a straightedge to approximate the line of best fit by hand.

c) Find an equation of the line of best fit for the data.

2. Mike is riding his bike home from his grandmother's house. In the table below, x represents the number of hours Mike has been biking and y represents the number of miles Mike is away from home. Make a scatter plot for this data on the grid below.

Hours (x)	1	2	3	4	5	6	7	8
Miles (y)	35	29	26	20	16	9	6	0



a) Describe the association between the data points on the scatter plot.

b) Use a straightedge to approximate the line of best fit.

c) Find an equation of the line of best fit for the data.

d) What does the slope represent in the context of the problem? What does the y-intercept represent in the context of the problem?

e) Could you use your equation to predict how far Mike would be after 10 hours? Use mathematics to justify your answer.

Line of Best Fit using the calculator

3) Use the table below to answer the questions about the population p (in millions) in Florida.

Year, t	2002	2003	2004	2005
Population (millions)	16.4	17.0	17.4	17.8

a) Find the best-fitting line for the data and the correlation coefficient.

b) Using this model, what will be the population in 2020?

---

4) Use the table below to answer the questions about the U.S. residential carbon dioxide emissions from 1993 to 2002. Emissions are measured in million metric tons.

Year, t	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Emissions	1027.6	1020.9	1026.5	1086.1	1077.5	1083.3	1107.1	1170.4	1163.3	1193.9

a) Find the best-fitting line for the data and the correlation coefficient.

b) Using this model, how many residential tons were emitted in 1990? In 2010?

---

5) Use the table below to answer the questions about the operating costs in thousands of a small business from 2000 to 2007.

Year, t	2000	2001	2002	2003	2004	2005	2006	2007
Operating Costs	2.3	2.6	3.1	3.3	4.0	5.2	5.9	7.0

a) Find the best-fitting line for the data and the correlation coefficient.

b) Using this model, what will be the operating costs in 2015?

Name: ______________________________ Date: __________ Period_________

Scatter Plots and Line of Best Fit – TV Task

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use the given functions or choose a function suggested by the context. Emphasize linear and exponential models.
MCC9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

1. Students in Ms. Garth's Algebra II class wanted to see if there are correlations between test scores and height and between test scores and time spent watching television. Before the students began collecting data, Ms. Garth asked them to predict what the data would reveal. Answer the following questions that Ms. Garth asked her class.

   a. Do you think students' heights will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation?

   b. Do you think the average number of hours students watch television per week will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation? Do watching TV and low test grades have a cause and effect relationship?

2. The students then created a table in which they recorded each student's height, average number of hours per week spent watching television (measured over a four-week period), and scores on two tests. Use the actual data collected by the students in Ms. Garth's class, as shown in the table below, to answer the following questions.

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Height (in inches)	60	65	51	76	66	72	59	58	70	67	65	71	58
TV hrs/week (average)	30	12	30	20	10	20	15	12	15	11	16	20	19
Test 1	60	80	65	85	100	78	75	95	75	90	90	80	75
Test 2	70	85	75	85	100	88	85	90	90	90	95	85	85

   a. Which pairs of variables seem to have a positive correlation? Explain.

   b. Which pairs of variables seem to have a negative correlation? Explain.

   c. Which pairs of variables seem to have no correlation? Explain.

3. Using the statistical functions of your graphing calculator, determine a line of good fit for each of the following categories.

   a. Score on Test 1 versus hours watching television:

   b. Score on test 1 versus score on test 2:

   c. Hours watching television versus score on test 2:

4. Use your answer to 3a to predict the test 1 score of someone who watches tv for 40 hours per week.

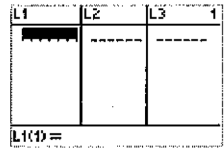
5. Use your answer to 3c to predict the test 2 score of someone who watches tv for 5 hours per week.

2. **If given a table of data:**

a. Perform a **LINEAR REGRESSION** using your calculator:

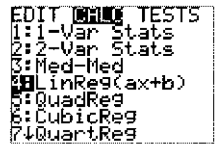
HOW TO PERFORM A LINEAR REGRESSION

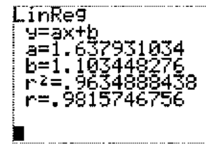
Enter Your Data List



> STAT
> 1: EDIT
> Type x values in L1
> Type y values in L2

Calculate the Linear Regression





> STAT
> CALC
> 4: LinReg(ax + b) … "a" is actually m (slope)
> substitute the "a" and "b" values into y = ax+ b to get your equation

Examples

x	y
5	1
10	2.5
15	4
20	6
25	7
30	8.5
35	11
40	12.5

Which equation defines the line of best fit for the data in the table?

F $y = \frac{1}{3}x - 10$ H $y = \frac{2}{3}x - 1$

G $y = \frac{1}{3}x - 1$ J $y = \frac{2}{3}x - 10$

Age of car (a)	Value of car (v)
0	12,500
1	9,200
2	7,850
4	6,100
8	3,425

Which equation most closely defines the line of best fit for the data?

A $v = a + 12{,}500$

B $v = 11{,}000a - 12{,}500$

C $v = {}^{-}1000a + 8{,}000$

D $v = {}^{-}1000a + 11{,}000$

Your Turn – Find the line of best fit

1. 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

a: ______________ b: ___________ Line of best fit: __________________________

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

x (in.)	60	62	64	66	68	70	72
y (lbs.)	105	114	124	131	139	149	158

a: ______________ b: ___________ Line of best fit: __________________________

Making Predictions

Once we have found our line of best fit, we can use that line to make predictions.

Using Your Turn #1 (hours spent studying vs. final exam grade):

a) Line of best fit: ______________________________

b) Predict the final exam grade for a student who studies:

i. 2.5 hours

ii. 6 hours

c) If a student earned a 98 on the exam, how many hours did he/she study?

Let's try another:

Using Your Turn #2 (sand sharks):

a) Line of best fit: ______________________________

b) Predict the weight of a sand shark with a length of:

i. 80 inches

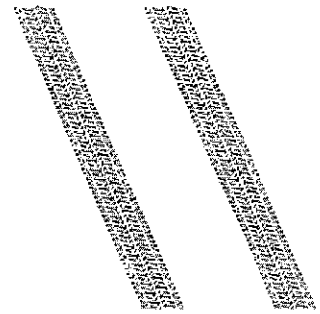
ii. 105 inches

c) If a sand shark weighs 250 pounds, what is its length?

Your Turn

Police investigating traffic accidents often estimate the speed of a vehicle by measuring the length of the tire skid distance. The following table gives the average tire skid distance for an automobile with good tires on dry pavement.

Tire Skid Distance (in feet)	Speed (in miles per hour)
25	28
54	35
89	45
132	55
184	65
244	75
313	85



1. Find the linear regression equation: ______________________.

2. Estimate the speed of a vehicle with the following tire skid mark distances:
   (are you finding x or y?)
   a) 150 feet

   b) 200 feet

3. Find the tire skid distance of a car travelling at the following speeds:
   (are you finding x or y?)
   a) 52 miles per hour

   b) 105 miles per hour

Example 2

Listed below are the yearly total enrollment (number of students) figures in Loudoun County Public Schools for the last 10 years.

Year	Enrollment
2005	47,361
2006	50,478
2007	54,047
2008	57,009
2009	60,096
2010	63,220
2011	65,668
2012	68,289
2013	70,858
2014	73,461

1. What is the linear regression equation for the data above?

2. Using your linear regression equation, what is the predicted enrollment for:

   a. 2015

   b. 2018

3. In what year can we expect the LCPS enrollment to surpass 95,000 students? (Hint: is "year" the x or the y?).

Name: ______________________________ Date: ________Period__________

Exponential Regression Homework

1) **Baseball Salaries**. Ball players have been signing ever-larger contracts. The highest salaries (in millions of dollars per season) for some notable players are given in the following table. **Remember to use 1980 as 0 and so on.**

Player	Year	Salary (millions $)
Nolan Ryan	1980	1.0
George Foster	1982	2.04
Kirby Puckett	1990	3.0
Jose Canseco	1990	4.7
Roger Clemens	1991	5.3
Ken Griffey, Jr.	1996	8.5
Albert Belle	1997	11.0
Pedro Martinez	1998	12.5
Mike Piazza	1999	12.5
Mo Vaughn	1999	13.3
Kevin Brown	1999	15.0
Carlos Delgado	2001	17.0
Alex Rodriguez	2001	25.2

a) Find an exponential model for this data. Use years since 1980 for the x variable.

b) What is the growth rate of salaries? State the growth rate in the context of the problem?

c) Predict a salary for the year 2012.

d) Using the Internet, see if you can find out the largest baseball salary this year? How does it compare to your prediction?

2) **Internet Users.** Reliable data about Internet use are hard to come by. But Nau Internet Surveys cites estimates of 18 million Internet users in the U.S. in 1995, 76 million users in 1998, and 119.2 million in 1999.

a) Using regression, determine whether a linear or an exponential function would be a better model? Explain. Write down both equations, rounding to 3 decimals.

b) What is the slope from the best-fit linear function? Interpret this in the context of the problem.

c) What was the annual growth factor and growth rate from the best-fit exponential function? Interpret the rate in the context of the problem.

d) Using the model that you felt was the best in part a, how many users do you predict for 2020?

3) **Coins.** A box containing 1,000 coins is shaken and the coins are emptied onto a table. Only the coins that land heads up are returned to the box, and then the process is repeated. The accompanying table shows the number of trials and the number of coins returned to the box after each trial.

a) Write an exponential regression equation, rounding the calculated values to the nearest ten-thousandth.

Trial	0	1	3	4	6
Coins Returned	1000	610	220	132	45

b) Use the equation to predict how many coins would be returned to the box after the eighth trial.

36

Name: ______________________________ Date: __________Period_________

Exponential Regression

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use the given functions or choose a function suggested by the context. Emphasize linear and exponential models.
MCC9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

Hot Coffee

The data at the right shows the cooling temperatures of a freshly brewed cup of coffee after it is poured from the brewing pot into a serving cup. The brewing pot temperature is approximately 180° F.

Time (mins)	Temp (F°)
0	179.5
5	168.7
8	158.1
11	149.2
15	141.7
18	134.6
22	125.4
25	123.5
30	116.3
34	113.2
38	109.1
42	105.7
45	102.2
50	100.5

a) Determine an exponential regression model equation to represent this data.

b) Decide whether the new equation is a "good fit" to represent this data.

c) Based upon the new equation, what was the initial temperature of the coffee? What is the decay rate?

d) Interpolate data: When is the coffee at a temperature of 106 degrees?

e) Extrapolate data: What is the predicted temperature of the coffee after 1 hour?

f) In 1992, a woman sued McDonald's for serving coffee at a temperature of 180° that caused her to be severely burned when the coffee spilled. An expert witness at the trial testified that liquids at 180° will cause a full thickness burn to human skin in two to seven seconds. It was stated that had the coffee been served at 155°, the liquid would have cooled and avoided the serious burns. The woman was awarded over 2.7 million dollars. As a result of this famous case, many restaurants now serve coffee at a temperature around 155°. How long should restaurants wait (after pouring the coffee from the pot) before serving coffee, to ensure that the coffee is not hotter than 155°?

g) If the temperature in the room is 76° F, what will happen to the temperature of the coffee, after being poured from the pot, over an extended period of time?

Practice Problems:

1. Estimates for world population vary, but the data in the accompanying table are reasonable estimates of the world population from 1800 to 2000.

Year	X	Total Population (millions)
1800		980
1850		1260
1900		1650
1950		2520
1970		3700
1980		4440
1990		5270
2000		6080

   a. Identify your independent and dependent variables.

   b. Generate a best fit exponential function using your variables. Round to 3 decimals.

   c. What does your model give for the growth rate? Describe this in the context of the problem.

   d. Using the function, estimate the world population in 1750 and 2050 to 3 decimal places.

---

2. **Town Planning:** The town planners designed their town for an optimal growth of 8% per year. The present school construction will serve a population of 200,000. Below is a table representing the growth from 1997 to 2003.

Year	X	Population
1997		50,000
1998		54,000
1999		58,000
2000		62,986
2001		68,024
2002		73,466
2003		79,344

   a) Find and write the model of a linear regression. Use the model to determine what the population was in 1977. Round to 2 decimals.

   b) Find and write the model of an exponential regression. Use the model to determine what the population was in 1977. Round to 2 decimals.

   c) Determine which model is better to use. Explain why you selected your model.

   d) Using the better model, predict what the population will be in the year 2017.

   e) In what year will the population double for the better model?

Performance Task: Equal Salaries for Equal Work?

Name________________________________ **Date**________________

The data table shows the annual median earnings for female and male workers in the United States from 1984 to 2004. Use the data table to complete the task. Answer all questions in depth to show your understanding of the standards.

Year	Women's median earnings (in dollars)	Men's median earnings (in dollars)
1984	8,675	17,026
1985	9,328	17,779
1986	10,016	18,782
1987	10,619	19,818
1988	11,096	20,612
1989	11,736	21,376
1990	12,250	21,522
1991	12,884	21,857
1992	13,527	21,903
1993	13,896	22,443
1994	14,323	23,656
1995	15,322	25,018
1996	16,028	25,785
1997	16,716	26,843
1998	17,716	28,755
1999	18,440	30,079
2000	20,267	30,951
2001	20,851	31,364
2002	21,429	31,647
2003	22,004	32,048
2004	22,256	32,483

Data provided by U.S. Census Bureau

1. Create two scatter plots, one for women's median earnings over time and one for men's median earnings over time. Describe two things you notice about the scatter plots.

2. Terry and Tomas are trying to decide what type of model will most accurately represent the data. Terry thinks that a linear model might be most appropriate for each scatter plot. Help Terry find reasonable linear function rules for each scatter plot. Explain how you found these.

3. Using the linear models, will women's annual median earnings ever equal those of men? Why or why not?

4. Tomas thinks that an exponential model might be most appropriate for each scatter plot. Help Tomas find reasonable exponential function rules for each scatter plot. Explain how you found these.

40

5. Using the exponential models, will women's annual median earnings ever equal those of men? Why or why not?

6. If you answered yes to either question 3 or question 5, use that model to determine the first year women will have higher median earnings than men. Explain how you found your answer.

7. For each year listed in the table, find the ratio of women's to men's annual median earnings expressed as a percentage. Use the data to create a scatter plot of percentage versus year. Based on this graph, do you think women's annual median earnings will ever equal those of men? Why or why not?

8. Considering the results of the scatter plot in question 7 above, do you think the linear model or exponential model makes more sense? Why?

Data on earnings by gender provided by:

U.S. Census Bureau. "Table P-41. Work Experience—All Workers by Median Earnings and Sex: 1967 to2005." *Historical Income Tables—People.* www.census.gov/hhes/www/income/histinc/p41ar.html. (Date retrieved: July 24, 2007.)

Algebra I | Unit 11: Comparing Functions | Practice

Day 7: Quadratic & Exponential Regression **Name:** ______________________

Practice Assignment

1. Write a quadratic function, in standard form, that passes through the given points.

   a. (-4, -1), (-2, 3), (-1, 8)

   b. (-2, 7), (-1, -6), (1, -20)

2. The following data table represents approximate heights for a ball thrown by a shot-putter as it travels x meters horizontally.

Distance (m)	Height (m)
7	8
20	15
33	24
47	26
60	24
67	21

   a. Find a quadratic model that represents the data.

   b. What would be the height of a ball that travels 80 meters?

   c. Is the ball in the air after 110 meters (according to your model)?

3. Gold's closing price each year is shown in the table.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Price ($/oz)	273	277	343	417	436	513	636	837	870	1088	1420

   a. Find a quadratic model for the gold price over time using x = 0 for 2000. Does the model provide a good fit for the data? Explain using your scatterplot and table on your calculator.

   b. What would be the price of gold in 2017 according to your model?

   c. Using Google, what is the actual closing price for gold as of yesterday? Is your model accurate for 2017?

4. A colony of bacteria grows exponentially. The table below shows the data collected daily.

a. Write an exponential regression equation for the data, rounding all values to the hundredths place.

Day (x)	Population (y)
0	200
1	425
2	570
3	800
4	1035
5	1650
6	2600

b. Explain what the "a" and "b" values represent in terms of the problem context.

c. What is the growth rate?

d. How many bacteria will be present on Day 10?

5. A cup of soup is left on a countertop to cool. The table below gives the temperatures, in degrees Fahrenheit, of the soup recorded over a 10-minute period.

a. Write an exponential regression equation for the data, rounding all values to the hundredths place.

Time in Minutes (x)	Temperature in °F (y)
0	180.2
2	165.8
4	146.3
6	135.4
8	127.7
10	110.5

b. Describe what the "a" and "b" values represents in terms of the problem context.

c. What is the decay rate?

d. If this trend continues, when will the soup reach 100 degrees Fahrenheit?

6. The table below gives the average annual cost (e.g. tuition, room, and board) for four year public universities. Let 1981 be t = 0.

a. Write an exponential regression equation for the data, rounding all values to the hundredths place.

Year	Average Annual Cost
1981	$2,550
1991	$5,243
2001	$8,653
2011	$15,918

b. Describe what the "a" and "b" values represents in terms of the problem context.

c. What is the growth rate?

d. If this trend continues, when will the average cost of attendance exceed $35,000?

7. The table below shows the amount of a decaying radioactive substance (in grams) that remained for selected years after 1990.

a. Write an exponential regression equation for the data, rounding all values to the hundredths place.

Years After 1990 (x)	0	2	5	9	14	17	19
Amount (y)	750	451	219	84	25	12	8

b. Describe what the "a" and "b" values represents in terms of the problem context.

c. What is the decay rate?

d. Using your equation, how much substance remained in 2002?

Coord. Algebra Support　　　　　　　　　　　　　　　　4.21 – Algebra Test Review

Name: ______________________　　　　　　　　　　　　　Date: __________

Use the following to review for you test. Work the Practice Problems on a separate sheet of paper.

What you need to know & be able to do	**Things to remember**	**Problem**	**Problem**
Identify the measures of central tendency.	• Mean • Median • Mode	**1.** 36, 39, 58, 42, 106, 39, 48, 45	**2.** 50, 55, 60, 58, 62, 57, 68, 51, 63
Identify the measures of spread.	• Q1 • Q3 • IQR • Minimum • Maximum • Range • MAD	**3.** (Use the same #s from 1)	**4.** (Use the same #s from 2)
Construct a box-and-whisker plot.	• First dot: Min • First Line: Q1 • Middle Line: Median • Third Line: Q3 • Last dot: Max • Outlier: Q1 – 1.5(IQR) Q3 + 1.5(IQR)	**5.** Using the data from #1 & 3, construct a box and whisker plot.  **6.** Are there any outliers? Show your work!	
Determine if the situation has a positive, negative, or no correlation and if there is causation.	• Positive: Both items are increasing/decreasing • Negative: one item increases as the other decreases • No Correlation: No relationship • Causation: One item causes the other.	**7.** Practicing Free Throws vs. Free Throw Percentage **9.** Weight vs. Amount of Exercise	**8.** Colors of the Sky vs. Time of Day **10.** Number of Followers on Twitter vs. Number of Friends on Facebook

45

Find the line of best fit.

- $y = ax + b$
- r = correlation coefficient (if close to 0 bad fit; if close to 1 or -1 good fit.)

11. Determine the line of best fit. Is this model a good fit for the data?

Price	4.00	5.50	3.50	8.00	5.50	7.00
# of Sandwiches	68	55	85	22	64	28

Find the exponential regression model.

- $y = a(b)^x$
- r = correlation coefficient (if close to 0 bad fit; if close to 1 or -1 then good fit.)

12. Determine the exponential regression model. Is this model a good fit for the data?

Year	0	2	4	7
Revenue	3	4	11	25

Construct a probability table.

- Joint Probability: Individual Cell/Table Total
- Marginal Probability: Row or Column Total/ Table Total
- Conditional Probability: Individual Cell/Row or Column Total

Complete the table to answer the following questions.

	Football	Basketball	Soccer	
Males	48	35	17	
Females	22	38	40	

13. What is the probability that a randomly chosen female likes soccer?

14. What is the probability that someone likes basketball?

15. Given that a person likes football, what is the probability they are male?

-

Name: ______________________________ Date: ________Period__________

1. Identify the Five-Number Summary number for the data of Johnny's test scores and draw the Box & Whisker plot.

92, 96, 97, 83, 92, 58, 93, 88, 77, 48, 65, 80, 71

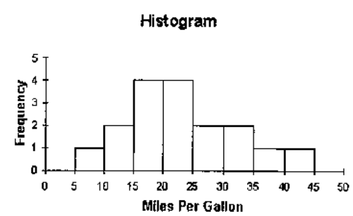
What is the range? ____ IQR? _____ MAD? _______

Are there any outliers in the data set?

2. The table gives the low temperatures in Chicago on eight randomly selected winter days. Which measure of central tendency probably gives the LEAST ACCURATE prediction of a "typical" low temperature on a Chicago winter day?

Chicago Lows							
17	25	28	12	16	55	18	22

3. Describe the shape of the distribution. Estimate the mean, median and upper and lower quartiles for the data.



4. Construct a frequency table from the following information:

A survey of 200 9th and 10th graders was given to determine what their favorite subject was. 72 said Math (50 which were freshmen), 38 said Social Studies (20 which were sophomores), and 40 freshmen and 50 sophomores said PE was their favorite.

				Total
Total				

Based on your table above, answer the following questions:

a) What are the marginal relative frequencies? ______

b) What are the joint relative frequencies? ______

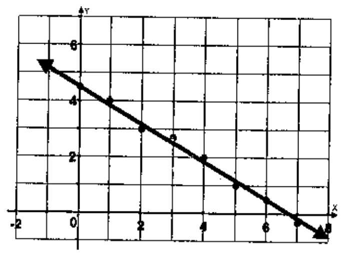
c) What is the probability that a student surveyed is a freshman? _____

d) What is the probability that a student surveyed likes Math? _____

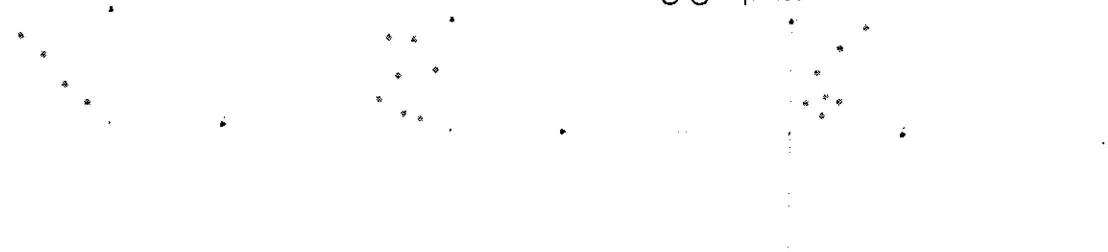
e) If a student likes Math, what is the probability that they are a freshman?_____

47

5. For the given scatter plot, find the equation for the line of best fit by hand.



6. Estimate the correlation coefficient for the following graphs.



7. Determine if the following situations represent a positive, negative, or no correlation.
   a) Number of hours studying for the SAT and your score. ___________
   b) The distance you drive and the number of stars in the sky. __________
   c) The temperature and the length of daylight hours for the day _________

8. Tell whether the following situations are causation: (yes or no)
   a) The number of boats on Lake Allatoona and the number of cars on the street ____
   b) The hours you work and the money you make ____
   c) The time spent studying and the A on the test _____

9. The following table shows a person study hours versus their test scores.

Hours studied (x)	2	5	1	0	4	2	3
Grade on test (y)	77	92	70	63	90	75	84

   a) Use your calculator to find the line of best fit for the data above. _______________
   b) What is the value of r? ______________________ Is this a good fit? _______________
   c) Use the equation to predict the test grade for someone who studies 5.5 hours. ___