mathematics vision project

Transforming Mathematics Education

SECONDARY MATH ONE

An Integrated Approach

MODULE 9

Modeling Data

MATHEMATICSVISIONPROJECT.ORG

The Mathematics Vision Project

Scott Hendrickson, Joleigh Honey, Barbara Kuehl, Travis Lemon, Janet Sutorius

MODULE 9 - TABLE OF CONTENTS

MODELING DATA

9.7 Getting Schooled – A Solidify Understanding Task

Use linear models of data and interpret the slope and intercept of regression lines with various units (S.ID.6, S.ID.7, S.ID.8)

READY, SET, GO Homework: Modeling Data 9.7

9.8 Rocking the Residuals – A Develop Understanding Task

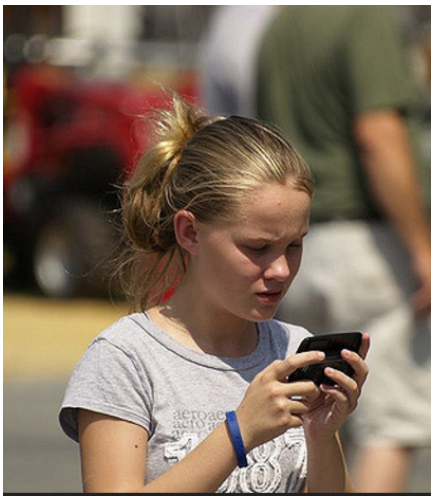
Use residual plots to analyze the strength of a linear model for data (S.ID.6)

READY, SET, GO Homework: Modeling Data 9.8

9.9 Lies and Statistics – A Practice Understanding Task

Use definitions and examples to explain understanding of correlation coefficients, residuals, and linear regressions (S.ID.6, S.ID.7, S.ID.8)

READY, SET, GO Homework: Modeling Data 9.9

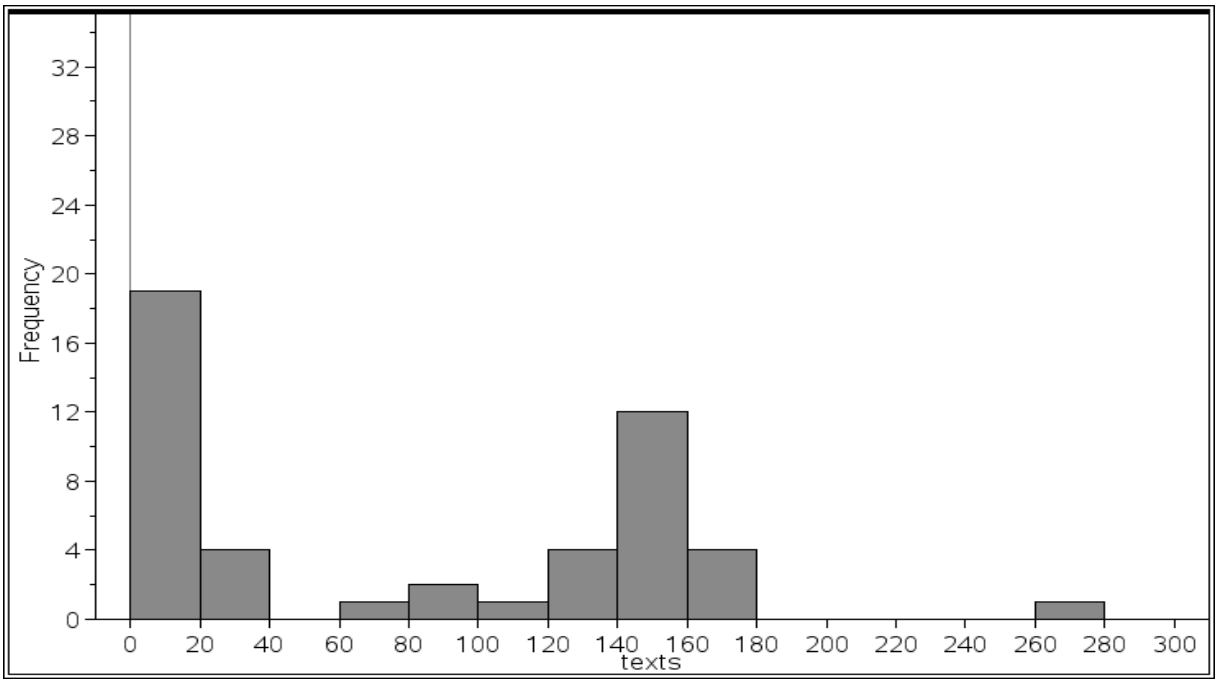


9.1 Texting by the Numbers

A Solidify Understanding Task

Technology changes quickly and yet has a large impact on our lives. Recently, Rachel was busy chatting with her friends via text message when her mom was trying to also have a conversation with her. Afterward, they had a discussion about what is an appropriate number of texts to send each day. Since they could not agree, they decided to collect data on the number of texts people send on any given day. They each asked 24 of their friends the following question: "What is the average number of texts you SEND each day?" The data and histogram representing all 48 responses:

{0, 2, 3, 3, 5, 5, 5, 5, 5, 5.5, 6, 6, 6, 10, 12, 13, 15, 15, 16, 20, 25, 35, 36, 70, 80, 85, 110, 130, 137, 138, 138, 140, 142, 143, 145, 150, 150, 150, 150, 150, 150, 150, 155, 162, 164, 165, 175, 275}



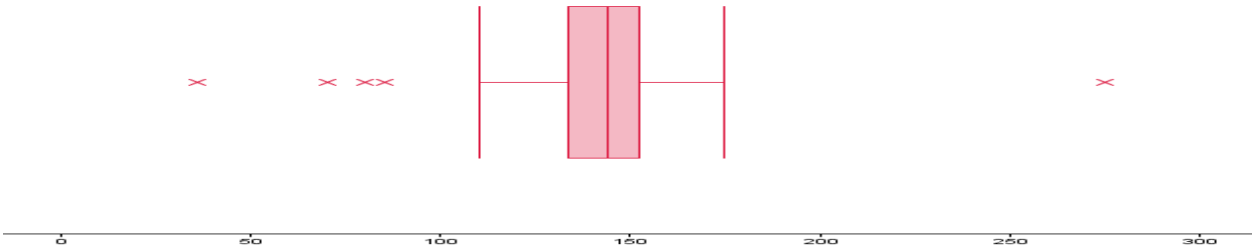
Part I:

1. What information can you conclude based on the histogram above?

2. Represent the same data by creating a box plot above the histogram.

3. What story does the box plot tell? Describe the pros and cons of each representation (histogram and box plot). In other words, what information does each representation highlight? What information does each representation hide or obscure?

Part II: Prior to talking about the data with her mom, Rachel had created a box plot using her own data she collected and it looked quite different than when they combined their data.



Average number of texts sent each day

4. Describe the data Rachel collected from her friends. What does this information tell you?

5. Compare the two box plots (Rachel's data vs all data).

6. Rachel wants to continue sending her normal number of texts (average of 100 per day) and her mom would like her to decrease this by half. Present an argument for each side, using mathematics to justify each person's request.

9.1 Texting by the Numbers – Teacher Notes

A Solidify Understanding Task

Purpose: In this task, students will use prior knowledge to interpret data using a histogram, and then represent the same data with a box plot. Students will discuss the shape (bimodal), center, and spread (outliers) of the data, the information highlighted or hidden by each representation, and compare two data sets using different representations. Comparing data sets is the focus of the task.

Core Standards Focus:

S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).
S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Related Standards: S.ID.2

Standards for Mathematical Practice of Focus in the Task:

SMP 1 – Make sense of problems and persevere in solving them

SMP 3 – Construct viable arguments and critique the reasoning of others

SMP 5 – Use appropriate tools strategically

The Teaching Cycle:

Launch (Whole Class):

Access background knowledge: Read the scenario from the task, then have students quietly write down their observations from the histogram, then share with a partner. Listen for comments to use during whole group discussion about shape, center, and spread. Have students move on to answer the remaining questions.

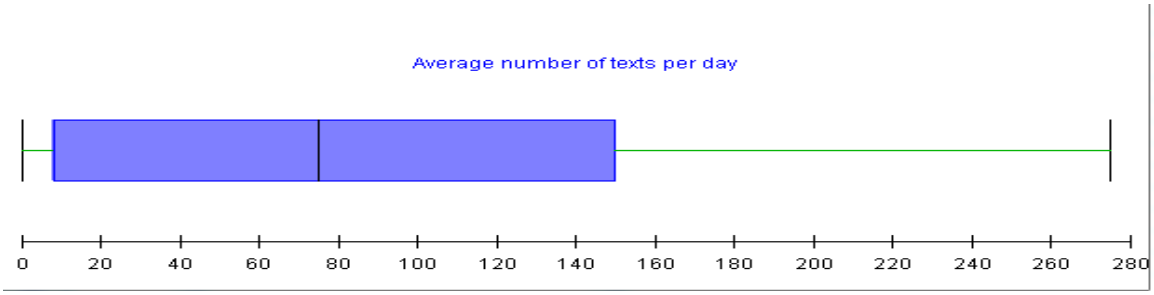
Note: If most students seem stuck, have the whole group come together to popcorn (quickly share) observations to get ideas about shape and spread out, then have students move on to answer the remaining questions related to the data in partners or small groups.

Explore (Small Group), part 1:

As you monitor, listen to the interpretation of the histogram and press students to describe the distribution (shape, center, and spread). Look for students who talk about the data having two 'modes' and their conjectures for why that may be (have them share the two mode conversation during the whole group discussion). When students are creating the box plot, remind them to label each quartile and listen for comments about the the data point of 275 (see if someone labels this as an outlier). Students have created box plots before. After most students have created the box plot to go with the same data and several have already written about the information each representation highlights, bring the class back together for the first whole group discussion.

Discuss (Whole Class), part I:

For this discussion, be sure to have the histogram displayed so everyone can make a visual connection to the description of the data. As the first student chosen shares their interpretation of the histogram, make sure they point to the histogram as they communicate their interpretation of the data. Students may not use the academic vocabulary of bimodal, but now is a good time to bring this up and have students write this in their journal. Next have a student share their box plot and go over their interpretation of the same data (quartiles, median, variability, and possibly outlier).



If no one in your class has a box plot that shows 275 as an outlier, then their box plot will look like the one above. If you have students who made 275 an outlier, have the outlier discussion first, then compare what each representation (histogram, box plot) highlights and what each representation 'hides' or obscures. If no one made 275 an outlier, then start with comparing representations then discuss what determines an outlier.

In summary, for this whole group discussion, be sure to get out the following:

- two modes with a description that people either seem to only do a few texts (less than 20) or a lot (between 140-180) per day,
- the data mostly lies between 0 and 180 texts per day; one value appears to be an outlier of 275 texts per day.
- the mean for the whole set of data is 81.05- does this seem relevant?
- Pros of histogram: shows frequency, can see bimodal data.
  Pros of box plot: shows quartile ranges and the median
- In this situation, the box plot obscures the bimodal data and may seem like there is a more even distribution.

If the outlier discussion hasn't happened, now is the time to talk about data points that seem to be outliers. A common rule of thumb to determine if a data point is an outlier is to use the equation: 1.5 (value of interquartile range)+ value of q3 for upper extremes or 1.5 (value of interquartile range) - value of q1 for lower extremes. Ask students if this data has any outliers? Because the interquartile range is so large, the value of 275 is not an outlier. Afterward, explain how outliers can be represented in a box plot and show what the box plot would look like if 275 had been an outlier.

Explore (Small Group), part II:

For part II, students should conclude that Rachel's friends are mostly representing the 'upper mode' of the data while her mom's friends are the 'lower mode'. Students may have already 'guessed' this during the first part of the task, however, they should now use the data from the box plot in part II to justify this statement.

If you would like your students to have the data after analyzing Rachel's box plot, it is listed here:

Rachel's mom: {150, 5.5, 6, 5, 3, 10, 150, 15, 20, 15, 6, 5, 3, 6, 0, 5, 12, 25, 16, 35, 5, 2, 13, 5}

Rachel: {130, 145, 155, 150, 162, 80, 140, 150, 165, 138, 175, 275, 85, 137, 110, 143, 138, 142, 164, 70, 150, 36, 150, 150}

Discuss (Whole Class), part II:

The focus of this discussion is initially on the interpretation of Rachel's data, then on the comparison of the two sets of data (Rachel's friends versus Rachel's moms friends).

Aligned Ready, Set, Go: Features 9.1

9.1

READY, SET, GO! **Name** **Period** **Date**

READY

Topic: Measures of central tendency

Sam's test scores for the term were 60, 89, 83, 99, 95, and 60.

1. Suppose that Sam's teacher decided to base the term grade on the mean.
   a. What grade would Sam receive?

   b. Do you think this is a fair grade? Explain your reasoning.

2. Suppose that Sam's teacher decided to base the term grade on his median score.
   a. What grade would Sam receive?

   b. Do you think this is a fair grade? Explain your reasoning.

3. Suppose that Sam's teacher decided to base the term grade on the mode score.
   a. What grade would Sam receive?

   b. Do you think this is a fair grade? Explain your reasoning.

4. Aiden's test scores for the same term were 30, 70, 90, 90, 91, and 99. Which measure of central tendency would Aiden want his teacher to base his grade on? Justify your thinking.

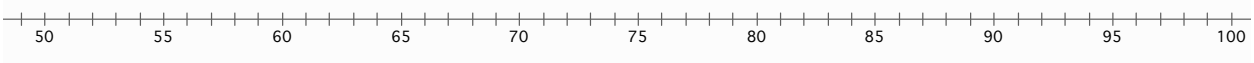
5. Most teachers base grades on the mean. Do you think this is a fair way to assign grades? Why or why not?

SET

Topic: Examining data distributions in a box-and-whisker plot.

6. Make a box-and-whisker plot for the following test scores.

60, 64, 68, 68, 72, 76, 76, 80, 80, 80, 84, 84, 84, 84, 88, 88, 88, 92, 92, 96, 96, 96, 96, 96, 96, 96, 100, 100



7 a. How much of the data is represented by the box?

b. How much is represented by each whisker?

8. What does the graph tell you about student success on the test?

GO

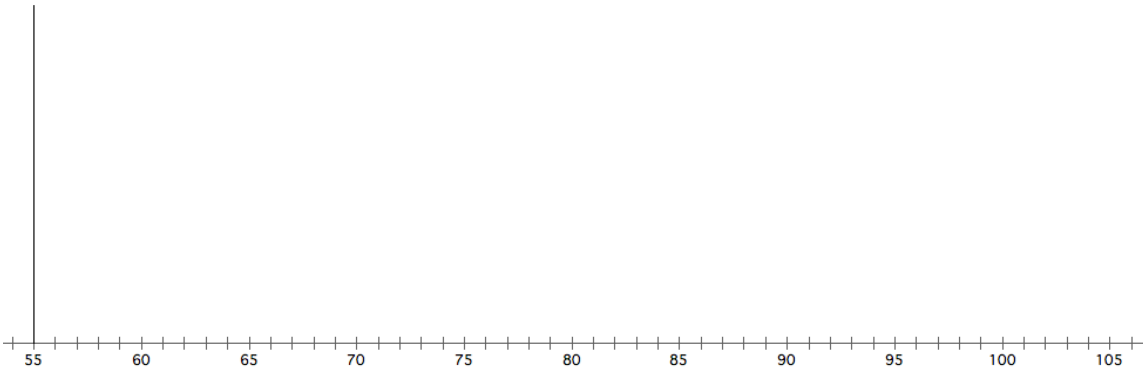
Topic: Creating histograms.

Use the data from the SET section to answer the following questions.

9. Make a frequency table with intervals. Use an interval of 5.

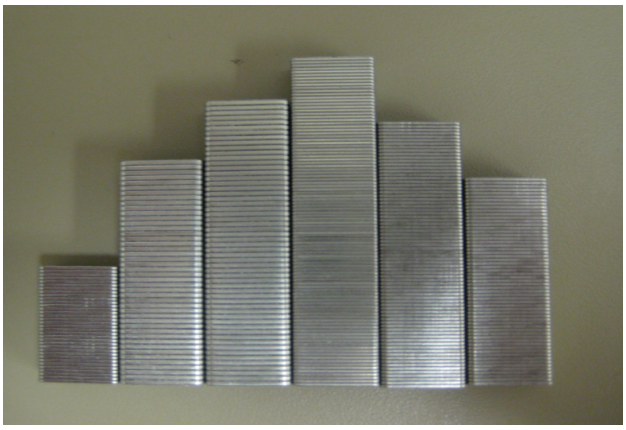
10. Make a histogram of the data using your intervals of 5.

Score	Frequency
60 – 64	
65 – 69	
70 – 74	
75 – 79	
80 – 84	
85 – 89	
90 – 94	
95 – 99	
100-104	



11. What information is highlighted in the histogram?

12. What information is highlighted in the box-and-whisker plot?



9.2 Data Distribution

A Practice Understanding Task

A lot of information can be obtained from looking at data plots and their distributions. It is important when describing data that we use context to communicate the **shape, center,** and **spread.**

Shape and spread:

- **Modes:** uniform (evenly spread- no obvious mode), unimodal (one main peak), bimodal (two main peaks), or multimodal (multiple locations where the data is relatively higher than others).
- **Skewed distribution:** when most data is to one side leaving the other with a 'tail'. Data is skewed to side of tail. (if tail is on left side of data, then it is skewed left).
- **Normal distribution and standard deviation:** curve is unimodal and symmetric. Data that has a normal distribution can also describe the data by how far it is from the mean using standard deviation.
- **Outliers:** values that stand away from the body of the distribution. For a box-and-whisker outliers determined if they are more than 1.5 times the interquartile range (length of box) beyond quartiles 1 and 3. Also considered an outliner if data is more than two standard deviations from the center of a normal distribution.
- **Variability:** values that are close together have low variability; values that are spread apart have high variability.

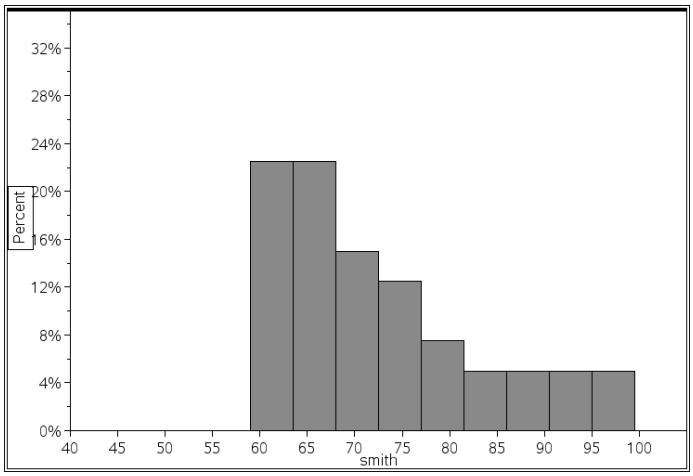
Center:

- Analyze the data and see if one value can be used to describe the data set. Normal distributions make this easy. If not a normal distribution, determine if there is a 'center' value that best describes the data. Bimodal or multimodal data may not have a center that would provide useful data.

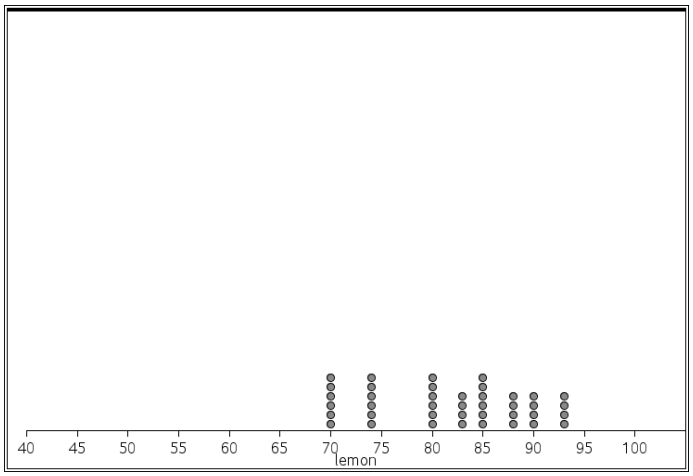
There are representations of test scores from six different classes found below, for each:

1. Describe the data distribution.
2. Compare data distributions between Anderson and Williams.
3. Compare data distributions between Williams and Lemon.
4. Compare data distributions between Croft and Hurlea.
5. Compare data distributions between Jones, Spencer, and Anderson.
6. Compare data distributions between Spencer and the other histograms.
7. Which distributions are most similar? Different? Explain your answer.

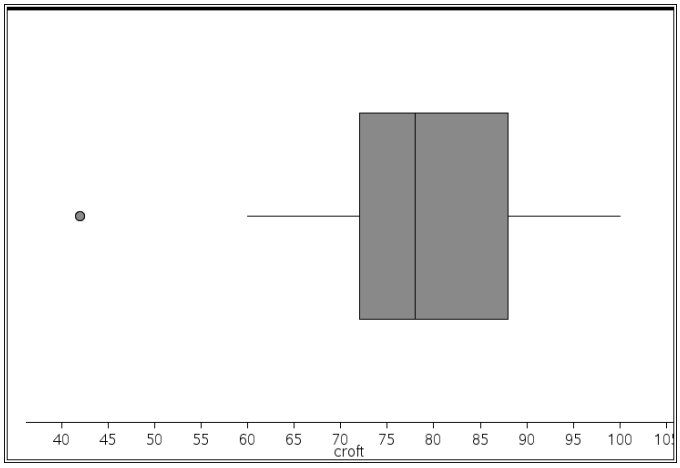
Data set I: Williams's class



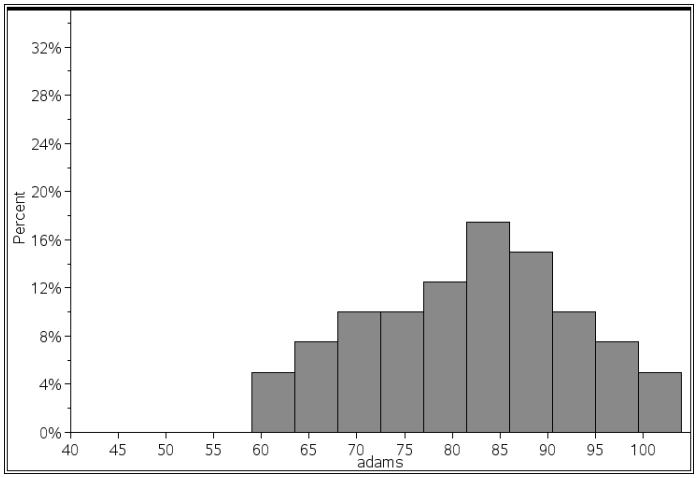
Data set II: Lemon's class



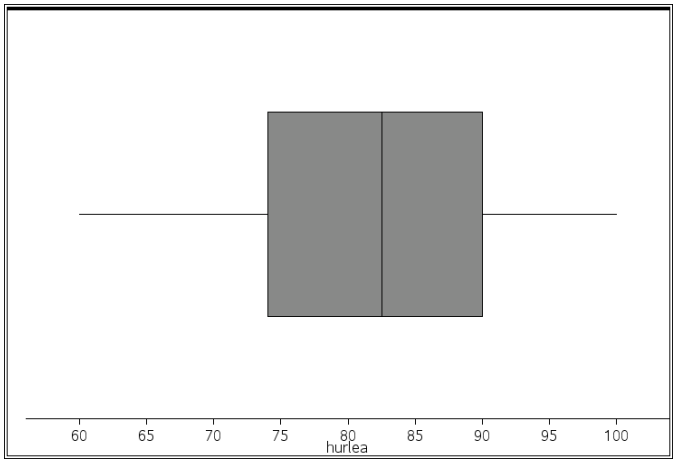
Data set III: Croft's Class



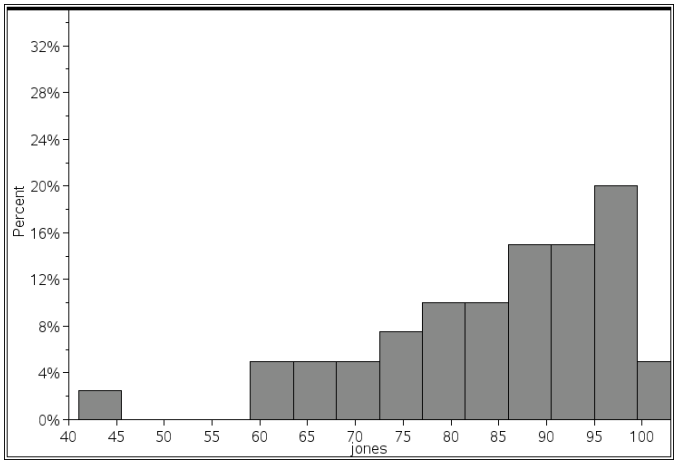
Data set IV: Anderson's Class



Data set V: Hurlea's class

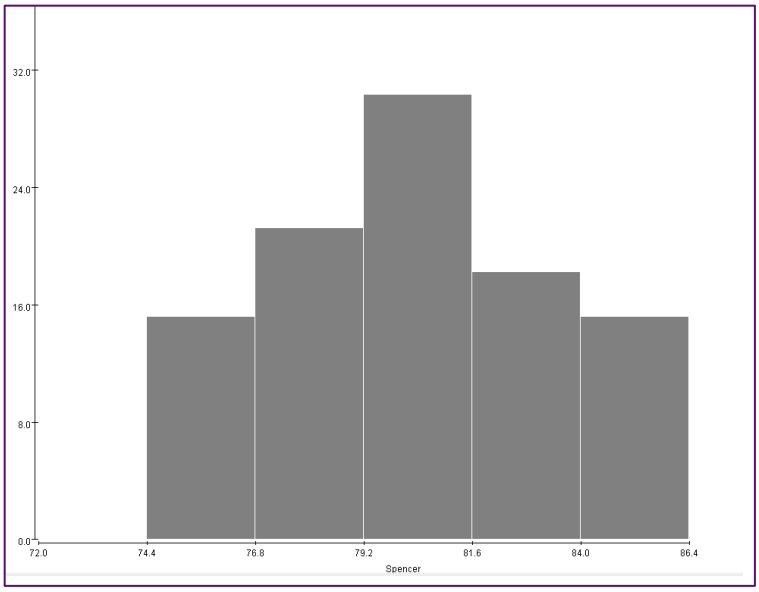


Data set VI: Jones' class

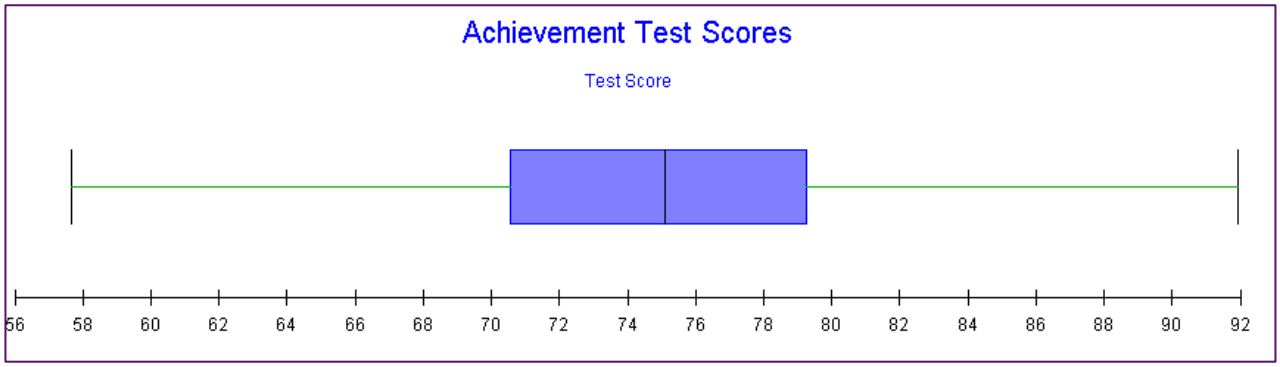


mathematics vision project

Data set VII: Spencer's class



Data set VIII: Overall Achievement Test Scores



9.2 Data Distribution – Teacher Notes

A Practice Understanding Task

Purpose: Students are already familiar with dot plots, box plots, and histograms. This task has them describe data distributions and compare shape, center, and spread of two or more sets of data.

Core Standards Focus:

> **S.ID.2** Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
>
> **S.ID.3** Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Related Standards: S.ID.1

Standards for Mathematical Practice:

> **SMP 3 – Construct viable arguments and critique the reasoning of others**
> **SMP 4 – Model with mathematics**
> **SMP 8 – Look for and express regularity in repeated reasoning**

The Teaching Cycle:

Note: It would be good to have the data you want to compare in a format that is large and visible for the whole group discussion. For example, you could copy the two data sets you wish to compare and place them next to each other in a format that can be projected so that when students are sharing during whole group, the visual representation is available for everyone to see.

Note: Students have been asked to identify and interpret univariate data using dot plots, histograms, and box plots since sixth grade. In this course, students are asked to **compare** data sets using their knowledge of shape, center, and spread and have become more comfortable with these attributes. Outliers, skewed data, and normal distribution may be new this year as well.

Launch (Whole Class):
Have students read the vocabulary to describe data distributions and ask them to underline information that is new to them. Have them work individually for a while on question 1 that has them describe each data set before having them work together with a partner or small group to answer the remaining questions (where they compare data sets).

Explore (Small Group):
Give students time to answer the questions comparing data sets. Listen for students to use vocabulary in describing a given data set, and to compare shape, center, and spread of two or more data sets. Listen for students to compare data sets, not just list attributes of each. Press students to make comparisons showing they understand when to use data to describe and compare shape, center, and spread between data sets. Examples include noticing outliers, variability and spread between data (notice that Hurlea and Spencer have a scale that is different than the others), and other trends. Again, make sure students do not just list characteristics of each distribution and think they are 'comparing'.

Discuss (Whole Class):
Begin the whole group discussion by selecting problems from questions that compare data sets. Based on small group conversations, choose which comparisons to share out in whole group. The focus of the whole group discussion is to do the following:

- Show student understanding of using statistics appropriate to the shape of the data distribution to compare center and spread.
- Show student understanding of what information is provided when given a histogram, box plot, dot plot.

Aligned Ready, Set, Go: Modeling Data 9.2

READY, SET, GO! **Name** **Period** **Date**

READY

Topic: Drawing conclusions from data.

In problems 1 – 4 you are to select the best answer based on the given data. Below your chosen answer is a confidence scale. Circle the statement that best describes your confidence in the correctness of the answer you chose. *The goal is to gain awareness of how it seems easier to draw conclusions in some cases than in others.*

1. Data: 1, 2, 4, 8, 16, 32, The next number in the list will be: _____

a. larger than 32 b. positive c. exactly 64 d. less than 32

I am certain I am correct. I am a little unsure. I had no idea so I guessed.

What about the data made you feel the way you did about the answer you marked?

2. Data: 47, -13, -8, 9, -23, 14, The next number in the list will be: _____

a. positive b. negative c. less than 100 d. less than -100

I am certain I am correct. I am a little unsure. I had no idea so I guessed.

What about the data made you feel the way you did about the answer you marked?

3. Data: -10, ¾, 38, -10, ½, -81, -10, ¼, 93, -10, The next number in the list will be: _____

a. more than 93 b. negative c. a fraction d. a whole number

I am certain I am correct. I am a little unsure. I had no idea so I guessed.

4. Data: 50, -43, 36, -29, 22, -15 The next number in the list will be: _____

a. odd b. less than 9 c. two-digits d. greater than -15

I am certain I am correct. I am a little unsure. I had no idea so I guessed.

What about the data made you feel the way you did about the answer you marked?

SET

Topic: Creating histograms.

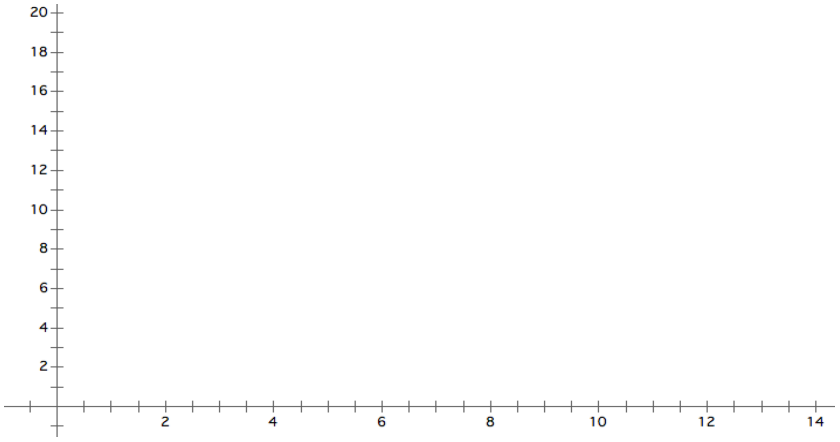
Mr. Austin gave a ten-point quiz to his 9th grade math classes. A total of 50 students took the quiz. Mr. Austin scored the quizzes and listed the scores alphabetically as follows.

1st Period Math	2nd Period Math	3rd Period Math
6, 4, 5, 7, 5, 9, 5, 4, 6, 6, 8, 5, 7, 5, 8, 1, 8, 7, 10, 9	4, 5, 8, 6, 8, 9, 5, 8, 5, 1, 5, 5, 7, 5, 7	9, 8, 10, 5, 9, 7, 8, 9, 8, 5, 8, 10, 8, 8, 5

5. Use ALL of the quiz data to make a frequency table with intervals. Use an interval of 2.

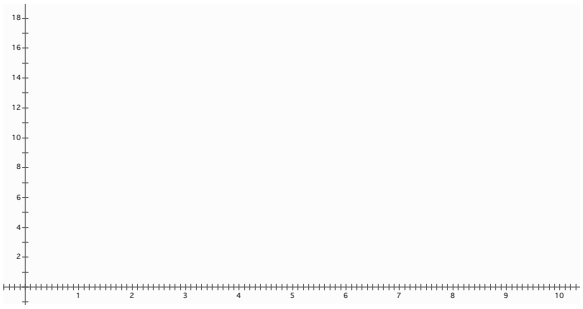
Score	Frequency
0 - 1	
2 – 3	
4 – 5	
6 – 7	
8 – 9	
10-11	

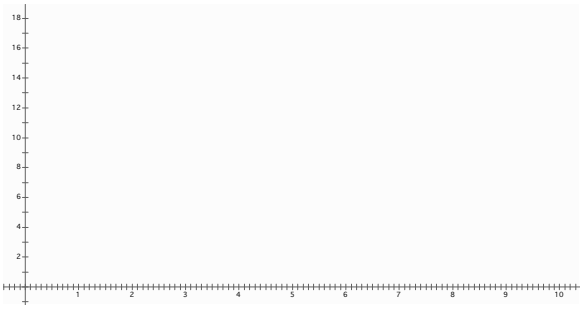
6. Use your frequency table to make a histogram for the data



7. Describe the data distribution of the histogram you created. Include words such as: *mode, skewed, outlier, normal, symmetric, center,* and *spread,* if they apply. (Hint: Don't forget standard deviation.)

8. Create a graph of your choice (histogram, boxplot, dotplot) for 1st and 3rd period.





9. Which class performed better? Justify your answer by comparing the shape, center, and spread of the two classes. (Hint: Don't forget standard deviation.)

GO

Topic: Figuring percentages

10. What percent of 97 is 11?

11. What percent of 88 is 132?

12. What percent of 84 is 9?

13. What percent of 88.6 is 70?

14. What is 270% of 60?

15. What is 84% of 25?

9.3 After School Activity

A Develop Understanding Task

Part I

Rashid is in charge of determining the upcoming after school activity. To determine the type of activity, Rashid asked several students whether they prefer to have a dance or play a game of soccer. As Rashid collected preferences, he organized the data in the following two-way frequency table:

	Girls	Boys	Total
Soccer	14	40	54
Dance	46	6	52
Total	60	46	106

Rashid is feeling unsure of the activity he should choose based on the data he has collected and is asking for help. To better understand how the data is displayed, it is useful to know that the outer numbers, located in the margins of the table, represent the total frequency for each row or column of corresponding values and are called *marginal frequencies* . Values that are part of the 'inner' body of the table are created by the intersection of information from the column and the row and they are called the *joint frequencies.*

1. Using the data in the table, construct a viable argument and explain to Rashid which after school event he should choose.

Part II: Two way frequency tables allow us to organize categorical data in order to draw conclusions. For each set of data below, create a frequency table. When each frequency table is complete, write three sentences about observations of the data, including any trends or associations in the data.

2. **Data set:** There are 45 total students who like to read books. Of those students, 12 of them like non-fiction and the rest like fiction. Four girls like non-fiction. Twenty boys like fiction.

	Fiction	Nonfiction	Total
Boys			
Girls			
Total			

Observation 1:
Observation 2:
Observation 3:

3. **Data set:** 35 seventh graders and 41 eighth graders completed a survey about the amount of time they spend on homework each night. 50 students said they spent more than an hour. 12 eighth graders said they spend less than an hour each night.

			Total
More than one hour			
Less than one hour			
Total			

Observation 1:
Observation 2:
Observation 3:

9.3 After School Activity – Teacher Notes

A Develop Understanding Task

Purpose: The purpose of this task is for students to make sense of two way frequency tables, to use the data to make an informed decision, and then construct a viable argument justifying their choice. Students will focus on different areas of the two way table so it is important that they are precise with their communication.

Core Standards Focus:
S.ID.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.

Standards for Mathematical Practice of Focus in the Task:
SMP 1 – Make sense of problems and persevere in solving them
SMP 3 – Construct viable arguments and critique the reasoning of others
SMP 6 – Attend to precision
The Teaching Cycle:

Launch (Whole Class):
Read the scenario and clarify how a two way frequency table is created. Explain to students that their job is to interpret the table, choose the after school activity that makes the most sense to them, and then provide mathematical reasoning that would convince Rashid to make the same selection.

Explore (Small Group):
Give students time to interpret the data, moving from group to group making sure they are using mathematics to make sense of the data (for example, showing that 14 out of 106 girls chose soccer means that only 14% of all girls would like soccer to be the chosen after school activity). As you monitor, listen for different groups to select opposite after school activities. Press students to be

very clear, using precise language to describe their mathematics. This will be important during the whole group discussion since the percentage for each situation varies depending on which 'total' students choose. This task is more about becoming familiar with how to find different percentages in a two way table and not about conditional probabilities. As students move to part II, help groups that struggle by asking "What are the two types of categorical data being compared?" or have them read one sentence only, then ask "Which cell of the table can be filled in based on this information?"

Discuss (Whole Class):

As a whole class, have two different groups share their recommendations for the after school activity. Have the first group share that selected the activity that was least chosen by the class. Ask the class if they have any questions for the group who presented, then ask the class if anyone who had chosen the other after school activity has changed their mind, and if so, explain why. Next, have a group share that chose the other activity. The purpose of this discussion is to highlight how to summarize data in a two way table, so be sure that the presenters communicate how they found each percentage presented and that all students can summarize a two way table. Move to part 2 of the task and have someone explain how they set up the two way table for one of the problems in part 2. As a whole class, summarize the process for filling in a two way table.

Aligned Ready, Set, Go: Modeling Data 9.3

READY, SET, GO!

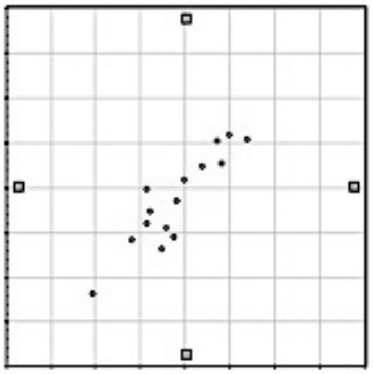
Name **Period** **Date**

READY

Topic: Interpreting data from a scatterplot

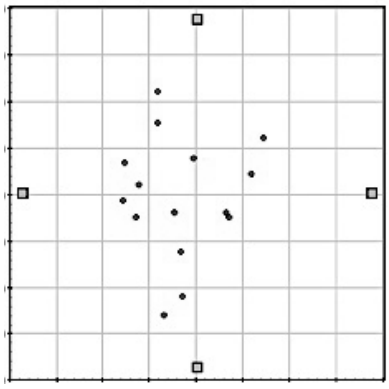
1. The scatter plot compares shoe size and height in adult males. Based on the graph, do you think there is a relationship between a man's shoe size and his height?

   Explain your answer.



2. The scatter plot compares left-handedness to birth weight. Based on the graph, do you think being left-handed is related to a person's birth weight?

   Explain your answer.



SET

Topic: Two-way frequency tables

Here is the data from Mr. Austin's 10-point quiz. Students needed to score 6 or better to pass the quiz.

1st Period Math	2nd Period Math	3rd Period Math
6, 4, 3, 7, 5, 9, 5, 4, 6, 6, 8, 5, 7, 3, 6, 2, 8, 7, 10, 9	3, 3, 8, 6, 6, 9, 5, 8, 5, 3, 5, 5, 7, 5, 7	9, 8, 10, 5, 9, 7, 8, 9, 8, 3, 8, 10, 8, 7, 5

3. Make a two-way frequency table showing how many students passed the quiz and how many students failed the quiz in each class.

	1st period	2nd period	3rd period	Total
Passed				
Failed				
Total				

Use a colored pencil to lightly shade the cells containing the *joint frequency* numbers in the table. The un-shaded numbers are the *marginal frequencies.* (Use these terms to answer the following questions.)

4. If Mr. Austin wanted to see how many students in all 3 classes combined passed the quiz, where would he look?

5. If Mr. Austin wanted to write a ratio of the number of passing students compared to the number of failing students for each class, where would he find the numbers he would need to do this?

6. Make a two-way frequency table that gives the *relative frequencies* of the quiz scores for each class.

	1st Period	2nd Period	3rd Period	Total
Passed				
Failed				
Total				

GO

Topic: Organizing data.

7. Sophie surveyed all of the 6th grade students at Reagan Elementary School to find out which TV Network was their favorite. She thought that it would be important to know whether the respondent was a boy or a girl so she recorded her information the following way.

Animal Planet	*Cartoon Network*	*Disney*	*Nickelodeon*
GGBBBB BGBBBGBBB GGBB BBBBBB	BBBBBBB BBGGGBBBG BGBGGGBGG	GGGGGGBBBBBB GBGBGG BBBGGBGG GGGBBBGGGGGB	BBBBGGGGGGGG GGGGGGBB GGGGBGGGGGGGGGBBBB BGGGGGGGG

Sophie planned to use her data to answer the following questions:

I. Are there more girls or boys in the 6th grade?
II. Which network was the boys' favorite?
III. Was there a network that was favored by more than 50% of one gender?

But when she looked at her chart, she realized that the data wasn't telling her what she wanted to know. Her teacher suggested that her data would be easier to analyze if she could organize it into a two-way frequency chart. Help Sophie out by putting the frequencies into the correct cells.

Favorite TV Networks	Girls	Boys	Totals
Animal Planet			
Cartoon Network			
Disney			
Nickelodeon			
Totals			

Now that Sophie has her data organized, use the two-way frequency chart to answer her 3 questions.

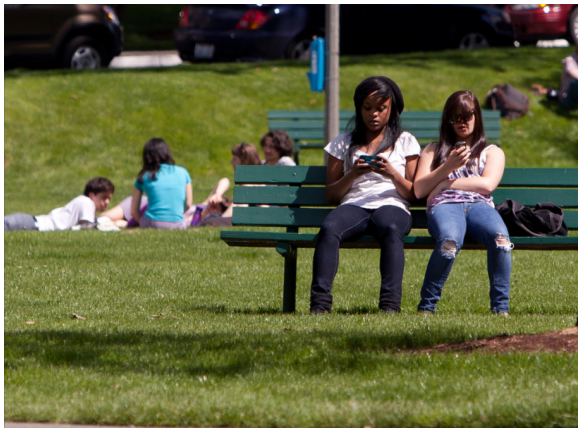
a. Are there more girls or boys in the 6th grade?

b. Which network was the boys' favorite?

c. Was there a network that was favored by more than 50% of one gender?

9.4 Relative Frequency

A Solidify/Practice Understanding Task



Rachel is thinking about the data she and her mom collected for the average number of texts a person sends each day and started thinking that perhaps a two-way table of the data they collected would help convince her mom that she does not send an excessive amount of texts for a teenager. The table separates each data point by age (teenager and adult) and by the average number of texts sent (more than 100 per day or less than 100 per day).

	Average is more than 100 texts sent per day	Average is less than 100 texts sent per day	Total
Teenager	20	4	24
Adult	2	22	24
Total	22	26	48

1. Write two observation statements of this two way table.

To further provide evidence, Rachel decided to do some research. She found that only 43% of people with phones send over 100 texts per day. She was disappointed that the data did not support her case and confused because it did not seem to match what she found in her survey.

2. What questions do these statistics raise for you? What data should Rachel look for to support her case?

After looking more closely at the data, Rachel found other percentages within the same data that seemed more accurate with the data she collected from her teenage friends.

3. How might Rachel use the data in the two way table to find percentages that would be useful for her case?

Part II: Once Rachel realized there are a lot of ways to look at a set of data in a two way table, she was motivated to learn about *relative frequency tables* and conditional frequencies. When the data is written as a percent, this is called a *relative frequency table*. In this situation, the 'inner' values represent a percent and are called conditional frequencies. The conditional values in a *relative frequency table* can be calculated as percentages of one of the following:

- the whole table (relative frequency of table)
- the rows (relative frequency of rows)
- the columns (relative frequency of column)

Since Rachel wants to emphasize that a person's age makes a difference in the number of texts sent, the first thing she decided to do is focus on the ROW of values so she could write conditional statements about the number of texts a person is likely to send based on their age. This is called a *relative frequency of row* table.

4. Fill in the percentage of teenagers for each of the conditional frequencies in the highlighted row below:

	Average is more than 100 texts sent per day	Average is less than 100 texts sent per day	Total
Teenager	20	4	24
% of teenagers	__ %	__%	100%
% of Adults	**2** 8%	**22** 92%	**24** 100%
% of People	**22** 46%	**26** 54%	**48** 100%

Row →

Since the PERCENTAGES created focus on ROW values, all conditional observations are specific to the information in the row. Complete the following sentence for the *relative frequency of row:*

5. Of all teenagers in the survey, _____ % average more than 100 texts per day.

6. Write another statement based on the *relative frequency of row:*

Below is the *relative frequency of column* using the same data. This time, all of the percentages are calculated using the data in the column.

	Average is more than 100 texts sent per day	Average is less than 100 texts sent per day	Total
Teenagers	**20** 91%	**4** 15%	**24** 50%
Adults	**2** 9%	**22** 85%	**24** 50%
Total	**22** 100%	**26** 100%	**48** 100%

7. Write two conditional statements using the *relative frequency of column*.

This data represents the *relative frequency of whole table*:

	Average is more than 100 texts sent per day	Average is less than 100 texts sent per day	Total
% of Teenagers	**20** 42%	**4** 8%	**24** 50%
% of Adults	**2** 4%	**22** 46%	**24** 50%
% of Total	**22** 46%	**26** 54%	**48** 100%

8. Create two conditional distribution statements for the *relative frequency of whole table.*

9. What information is highlighted when data is interpreted from *relative frequency tables*?

9.4 Relative Frequency – Teacher Notes

A Solidify Understanding Task

Purpose: In this task students will examine different ways to interpret *relative frequency tables* and will write conditional distribution statements based on the relative frequency of row, column, or whole table. Using data from *Texting By the Numbers,* students will see how two way tables can show information that is often hidden in box plots or histograms. They will also learn how conditional frequencies can provide specific information about a subgroup of the data (calling for more precision of describing the data).

Core Standards Focus:
S.ID.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.
Related Standards: S.ID.1, S.ID.2. S.ID.3

Standards for Mathematical Practice of Focus in the Task:
SMP 2 – Reason abstractly and quantitatively
SMP 6 – Attend to precision
SMP 7 – Look and make use of structure

The Teaching Cycle:
Launch (Whole Class):
As part of accessing background knowledge, you may wish to ask students what they remember about the data from the task *Texting By the Numbers.* The purpose is only to have students mention data from Rachel and her mom, with comments related to the story the data told (specifics not needed). Read the scenario from this task and have students answer the first question by writing and sharing a few observations about the two way table (review from the task *After School Activity*).

As statements are shared, write them on strips of paper (can be sorted during whole group discussion).

Explore (Small Group):

As students work through the task, listen for their conjectures about the data Rachel should focus on to make her case. After a few minutes, bring the class back together to discuss the types of *relative frequency tables.* Explain how each value is determined in the relative frequency of row. Have students work in pairs to complete the sentence frames and write conditional distribution statements for each of the three relative frequency tables. Give students time to consider all three tables and create statements about each table. Listen for understanding of each relative frequency table. To assist in writing sentences, remind students to pay attention to the focus of the table (whether the focus is the row, the column, or the entire table).

Discuss (Whole Class):

The intention of the whole group discussion is to highlight the following:

- differences between row, column, and whole table relative frequency statements
- become precise in our language as we use conditional frequency statements
- tell a story using two way tables.

One way to orchestrate this discussion is to select a student to share a specific relative frequency of column statement they have created (let them know this during the explore phase of the task) and have the class determine if the statement is from the relative frequency of row, column, or whole table. Then ask a student to share another relative frequency of column statement. Ask the group, what does the data specific to column tell us?

Move to showing the relative frequency of row statements and ask what does the data specific to row tell us? Continue discussion with relative frequency of whole table.

To conclude, discuss the last question from the task: What information is highlighted when data is interpreted from *relative frequency tables*? If there is time, also discuss how two way tables compare to other univariate models we have used (dot plots, box plots, histograms).

Aligned Ready, Set, Go: Modeling Data 9.4

READY, SET, GO! **Name** **Period** **Date**

READY

Topic: Writing explicit function rules for linear relationships

Write the explicit linear function for the given information below.

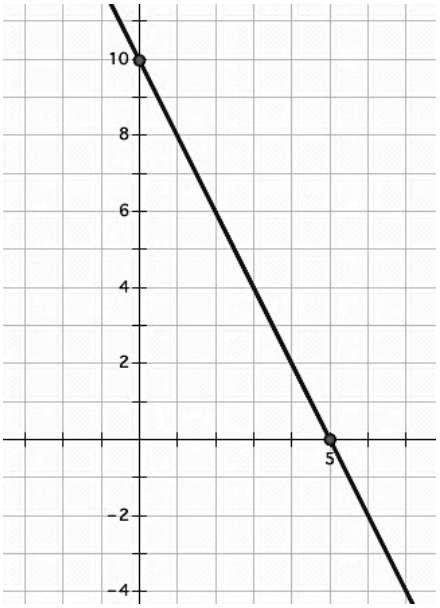
1. (3, 7) (5, 13)

2. Mike earns $11.50 an hour

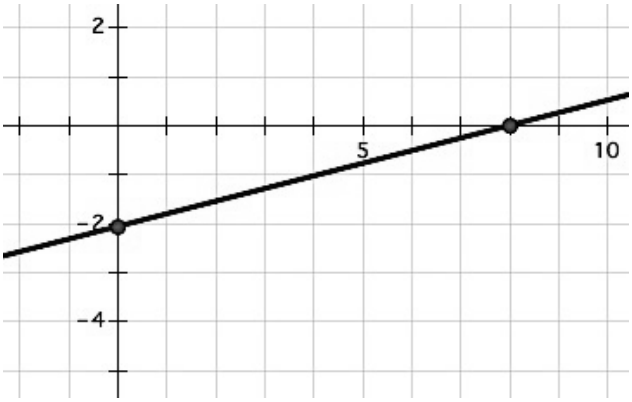
3. (-5, -2) (1, 10)

4. (-2, 12) (6, 8)

5.



6.



SET

Topic: Relative Frequency tables

For each two-way table below, create the indicated relative frequency table and also provide two observations with regard to the data.

7. This table represents survey results from a sample of students regarding mode of transportation to and from school.

	Walk	Bike	Car Pool	Bus	Total
Boys	37	47	27	122	233
Girls	38	22	53	79	192
Total	75	69	80	201	425

Create the *relative frequency of column table*. Then provide two observation statements.

	Walk	Bike	Car Pool	Bus	Total
Boys					
Girls					
Total	100%	100%	100%	100%	100%

8. The two-way table contains survey data regarding family size and pet ownership.

	No Pets	Own one Pet	More than one pet	Total
Families of 4 or less	35	52	85	172
Families of 5 or more	15	18	10	43
Total	50	70	95	215

Create the *relative frequency of row table*. Then provide two observation statements.

	No Pets	Own one Pet	More than one pet	Total
Families of 4 or less				100%
Families of 5 or more				100%
Total				100%

9. The two-way table below contains survey data about boys and girls shoes.

	Athletic shoes	Boots	Dress Shoe	Total
Girls	21	35	60	116
Boys	50	16	10	76
Total	71	51	70	192

Create the *relative frequency of whole table*. Then provide two observation statements.

	Athletic shoes	Boots	Dress Shoe	Total
Girls				
Boys				
Total				100%

GO

Topic: One variable statistical measures and comparisons

For each set of data determine the mean, median, mode, range, and standard deviation. Then create either a box-and-whisker plot or a histogram.

10. 23, 24, 25, 20, 25, 29, 24, 25, 30

11. 20, 24, 10, 35, 25, 29, 24, 25, 33

12. How do the data sets in problems 10 and 11 compare to one another?

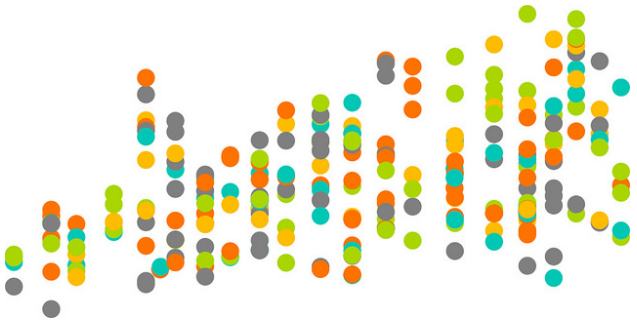
13. 2, 3, 4, 5, 3, 4, 7, 4, 4

14. 1, 1, 3, 5, 5, 10, 5, 1, 14

15. How do the data sets in problems 13 and 14 compare to one another?

9.5 Connect the Dots

A Develop Understanding Task



For each set of data:

- Graph on a scatter plot.
- Use technology (graphing calculator or computer) to calculate the correlation coefficient.

Set A

2	2.3	3.3	3.7	4.2	4.6	4.5	5	5.5	5.7	6.1	6.4
1	1.5	2.5	1.9	2.8	3.2	4.5	3.7	1.7	4.8	2.7	2.3

Set B

2	2.3	3.3	3.7	4.2	4.6	4.5	5	5.5	5.7	6.1	6.4
1	1.5	2.5	1.9	2.8	3.2	4.5	3.7	4	4.8	5	4.6

Set C

2	2.3	3.3	3.7	4.2	4.6	4.5	5	5.5	5.7	6.1	6.4
4.7	4.9	4.2	3.9	3.5	3.2	3.1	2.6	3.2	2.1	1.3	0.8

Set D

2	2.3	3.3	3.7	4.2	4.6	4.5	5	5.5	5.7	6.1	6.4
4.7	4.9	3.6	3.9	2.1	4.5	3.1	1.7	3.7	2.1	1.3	1.8

Set E

2	2.3	3.3	3.7	4.2	4.6	4.5	5	5.5	5.7	6.1	6.4
4.7	4	4.2	3.9	2.8	3.2	4.5	3.7	3.2	4.8	5	4.4

Set F

2	2.3	3.3	3.7	4.2	4.6	4.5	5
1.8	2.22	3.62	4.18	4.88	5.44	5.3	6

Set G

2	2.3	3.3	3.7	4.2	4.6	4.5	5
4.4	4.01	2.71	2.19	1.54	1.02	1.15	0.5

1. Put the scatter plots in order based upon the correlation coefficients.

2. Compare each scatter plot with its correlation coefficient. What patterns do you see?

3. Use the data in Set A as a starting point. Keeping the same *x*-values, modify the y-values to obtain a correlation coefficient as close to 0.75 as you can.

   Record your data here:

2	2.3	3.3	3.7	4.2	4.6	4.5	5	5.5	5.7	6.1	6.4

   What did you have to do with the data to get a greater correlation coefficient?

4. This time, again start with the data in Set A. Keep the same x-values, but this time, modify the y values to obtain a correlation coefficient as close to 0.25 as you can.

   Record your data here:

2	2.3	3.3	3.7	4.2	4.6	4.5	5	5.5	5.7	6.1	6.4

   What did you have to do with the data to get a correlation coefficient that is closer to 0?

5. One more time: start with the data in Set A. Keep the same x-values, modify the y-values to obtain a correlation coefficient as close to -0.5 as you can.

   Record your data here:

2	2.3	3.3	3.7	4.2	4.6	4.5	5	5.5	5.7	6.1	6.4

   What did you have to do with the data to get a correlation coefficient that is negative?

6. What aspects of the data does the correlation coefficient appear to describe?

7. On the night before the last math test, Shaniqua held a study group at her house. It was a great night; they ate a lot of pizza, did math, and laughed a lot. Shaniqua scored better on her test than usual and thought it might be related to pizza. She collected the following data from her friends in the study group:

	Shaniqua	David	Susana	Ruby	Deion	Oscar
Number of Pizza Slices Eaten	2	6	1	4	3	5
Increase in Test Score	5	9	4	7	6	8

Create a scatter plot of this data and calculate the correlation coefficient.

Based on these data, would you recommend eating pizza on the night before a test to increase scores? Why or why not?

8. Describe a situation with two variables that may have a high correlation, but not be causally related.

9. What are some reasons that two variables may be highly correlated but not have a causal relationship?

9.5 Connect the Dots – Teacher Notes

A Develop Understanding Task

Special Note to Teachers: This task requires the use of technology that can calculate the correlation coefficient, r. Most graphing calculators will work well. Free computer apps would be very helpful and easy to use on this task as well (GeoGebra and Desmos, etc.).

Purpose: The purpose of this task is to develop an understanding of the correlation coefficient. The task asks students to plot various data sets and use technology to calculate the correlation coefficient. They will order the graphs and create new data sets to develop the idea that the correlation coefficient indicates the strength and direction of a linear relationship in the data. Students also consider situations in which two variables are highly correlated, but the relationship is not necessarily causal.

Core Standards Focus:

S-ID. 8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

S-ID.9 Distinguish between correlation and causation.

S.ID Notes: Build on students' work with linear relationships in eighth grade and introduce the correlation coefficient. The focus here is on the computation and interpretation of the correlation coefficient as a measure of how well the data fit the relationship. The important distinction between a statistical relationship and a cause-and-effect relationship arises in S.ID.9.

Related Standards: S-ID.6

Standards For Mathematical Practice of Focus in the Task:

SMP-1 Make sense of problems and persevere in solving them.

SMP-5 Use appropriate tools strategically.

The Teaching Cycle

Launch (Whole Class): Since this is the first task in the module that uses scatter plots for bivariate data, begin by reminding students of the term and how they are constructed. Tell them

that the purpose of this task is to come up with their own hypothesis about what features of the data the correlation coefficient describes. Show students how to enter data and calculate a correlation coefficient using whatever technology you have selected for your class. Also tell students that correlation coefficients can only be calculated for two quantitative variables. For the purpose of the classroom discussion, each student should plot the data and record the correlation coefficient on paper (either by hand or using a printer) for each problem. This will facilitate comparing and ordering of the graphs, as well as use the data from problems 4-7 to confirm their hypothesis about the correlation coefficient.

Explore (Small Group): Monitor students as they work to see that they are able to use the technology properly and are recording the graphs on paper. Once they have graphed and ordered each of the first 6 data sets, encourage them to speculate and share their ideas in the groups. Listen for students that are noticing that the values of r are between -1 and 1, that negative values describe decreasing trends, positive values describe increasing trends, that values of r near 0 correspond to data without noticeable patterns, and r values near 1 or -1 describe data that appear to fit a linear model.

Discuss (Whole Class): Prepare for the discussion by reproducing the scatter plots for sets A-G (given below) so that they can be displayed for the entire class. Post the plots in order from -1 to 1 (G, C, D, E, A, B, F). Ask students for their ideas about the aspects of the data described by the correlation coefficient. Record a list that should include some or all of the following:

- r values range between -1 and 1,
- negative values of r describe negative association,
- positive values of r describe positive association,
- values of r near 0 correspond to data with a very weak linear relationship
- r values near 1 or -1 describe data that fit a linear model

They may also have some ideas that may be abandoned later, based upon the discussion.

Turn the discussion to #4. Ask three students to display the scatter plots they made that have a correlation coefficient of .75. Ask the class to see what the three graphs have in common, emphasizing observations about the direction of the association and the appearance of linearity. Ask students how they adjusted the data in set A, for which r = 0.49, to increase r. Ask the class how

the experience in # 4 either confirms or denies their hypotheses about the correlation coefficient. Move through questions 5 and 6 in similar fashion.

Tell students that a correlation of 1 or -1 is a perfect correlation, and ask what that means for the relationship between the two variables. Discuss the conclusions that they drew in question 7.

End the discussion by eliminating any remaining incorrect statements in the list of student ideas about the correlation coefficient and by writing and discussing the meaning of the following statement: ***The correlation coefficient measures the strength and direction of a linear relationship between two quantitative variables.***

Aligned Ready, Set, Go: ***Modeling Data 9.5***

READY, SET, GO! **Name** **Period** **Date**

READY

Topic: Estimating the line of best fit

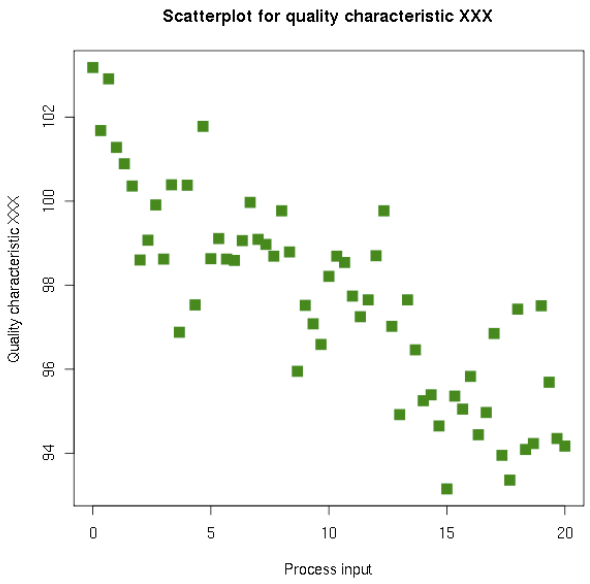
Examine the scatterplot below. Imagine that you drew a straight line through the general pattern of the points, keeping as close as possible to all points with as many points above the line as below.

1. Predict a possible y-intercept and slope for the line you imagined.

   a. y-intercept: _____

   b. slope: _____

2. Sketch the line that you imagined for question #1 and write an equation for that line.



SET

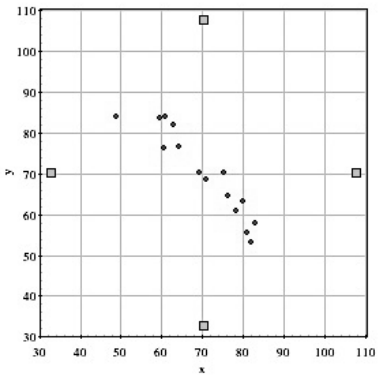
Topic: Estimating the correlation coefficient

Match the following scatterplots with the correct correlation coefficient.

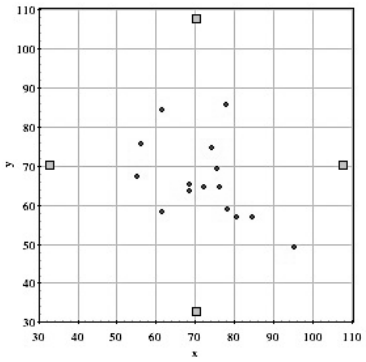
Possible correlation coefficients:

a. 0.05 b. 0.97 c. -0.94 d. -0.49 e. 0.68 f. -0.25

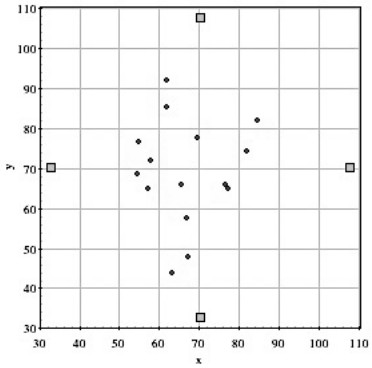
3.



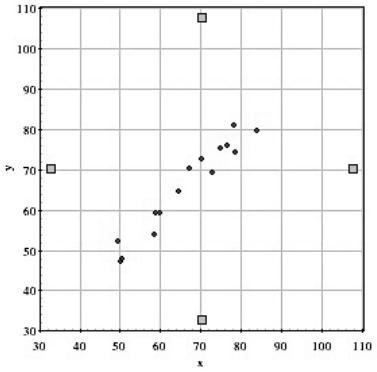
4.



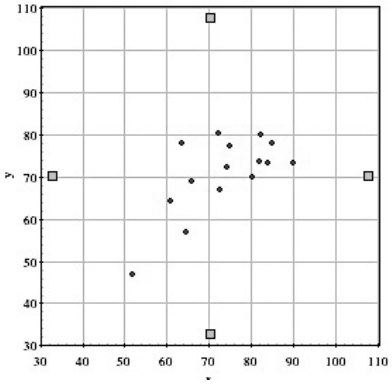
5.



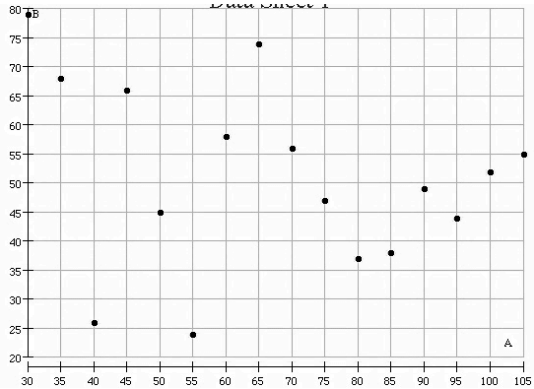
6.



7.



8.

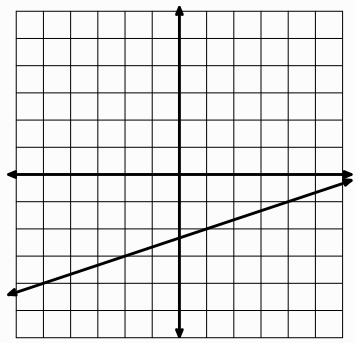


GO

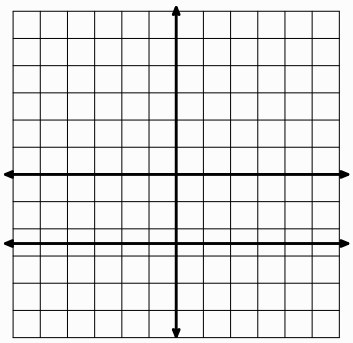
Topic: Visually comparing slopes of lines.

Follow the prompt to sketch the graph of a line on the same grid with the given characteristics.

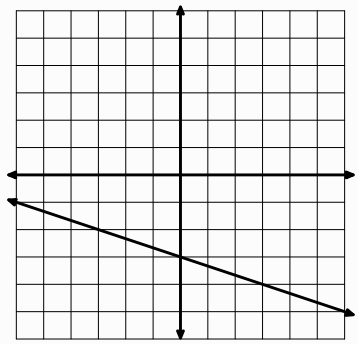
8. A greater slope



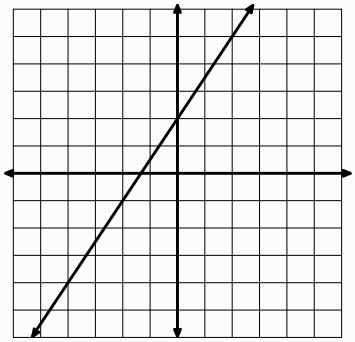
9. A lesser slope



10. A larger y–intercept and a lesser slope



11. Slope is the opposite reciprocal.



9.6 Making More $

A Solidify Understanding Task

Each year the U.S. Census Bureau provides income statistics for the United States. In the years from 1990 to 2005, they provided the data in the tables below. (All dollar amounts have been adjusted for the rate of inflation so that they are comparable from year-to-year.)

Year	Median Income for All Men
2005	41196
2004	41464
2003	40987
2002	40595
2001	41280
2000	41996
1999	42580
1998	42240
1997	40406
1996	38894
1995	38607
1994	38215
1993	37712
1992	37528
1991	38145

Year	Median Income for All Women
2005	23970
2004	23989
2003	24065
2002	23710
2001	23564
2000	23551
1999	22977
1998	22403
1997	21759
1996	20957
1995	20253
1994	19158
1993	18751
1992	18725
1991	18649

1. Create a scatter plot of the data for men, setting 1991 as year 1.

   What is your estimate of the correlation coefficient for these data?

2. On a separate graph, create a scatter plot of the data for women, setting 1991 as year 1.

   What is your estimate of the correlation coefficient for these data?

3. Estimate and draw lines that model each set of data.

4. Describe how you estimated the line for men. If you chose to run the line directly through any particular points, describe why you selected them.

5. Describe how you estimated the line for women. If you chose to run the line directly through any particular points, describe why you selected them.

6. Write the equation for each of the two lines in slope intercept form.

   a. Equation for men:

   b. Equation for women:

7. Use technology to find the actual correlation coefficient for men.

   What does it tell you about the relationship between income and years for men?

8. What is the actual correlation coefficient for women?

   a. What does it tell you about the relationship between income and years for women?

   b. What do the correlation coefficients for men and women tell us about how the data compares?

9. Use technology to calculate a linear regression for each set of data. Add the regression lines to your scatter plots.

c. Linear regression equation for men:

d. Linear regression equation for women:

10. Compare your model to the regression line for men. What does the slope mean in each case? (Include units in your answer.)

11. Compare your model to the regression line for women. What does the y-intercept mean in each case? (Include units in your answer.)

12. Compare the regression lines for men and women. What do the lines tell us about the income of men vs women in the years from 1991-2005?

13. What do you estimate will be the median income for men and women in 2015?

14. The Census Bureau provided the following statistics for the years from 2006-2011.

Year	Median Income for All Men
2011	37653
2010	38014
2009	38588
2008	39134
2007	41033
2006	41103

Year	Median Income for All Women
2011	23395
2010	23657
2009	24284
2008	23967
2007	25005
2006	24429

With the addition of these data, what would you now estimate the median income of men in 2015 to be? Why?

15. How appropriate is a linear model for men's and women's income from 1991-2011? Justify your answer.

9.6 Making More $ – Teacher Notes

A Solidify Understanding Task

Special Note to Teachers: This task requires the use of technology that can calculate the correlation coefficient, r. Most graphing calculators will work well. Free computer apps would be very helpful and easy to use on this task as well (GeoGebra and Desmos, etc.).

Purpose: The purpose of this task is to solidify understanding of correlation coefficient and to develop linear models for data. Students are asked to estimate and calculate correlation coefficients. In the task they estimate lines of best fit and then compare them to the calculated linear regression. The task demonstrates the dangers of using a linear model to extrapolate well beyond the actual data. The task ends with an opportunity to use the correlation coefficient and scatter plot to determine the appropriateness of a linear model.

Core Standards Focus:

S-ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

S-ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

Related Standards: S.ID.6

Standards For Mathematical Practice of Focus in the Task:

SMP 2 – Reason abstractly and quantitatively.

SMP 4 – Model with mathematics.

The Teaching Cycle

Launch (Whole Class): Introduce the task telling students that this task extends what they have done in previous modules to model situations with lines. In this case, they will be modeling real

data, which is not usually perfectly linear (correlation coefficient of 1 or -1). Before actually beginning the task of making scatter plots, ask students to make some observations about the data in the two tables that show the median income for men and women. They may notice that women's salaries are lower than men's or that they both appear to be increasing over time. What questions are raised by these observations? Ask students to work on questions 1-4.

Explore (Small Group): Monitor students as they are working, observing their thinking about the plots. Encourage them to discuss the correlation coefficient with their group, noticing both the direction and the strength of the linear relationship. Many students may not feel that a linear model is appropriate for the men's data because of the shape of the distribution. (Both scatter plots are shown below). Listen as students talk about their strategies for placing the line of best fit on the two scatter plots and be prepared to call on students with different strategies for the discussion. Some strategies that can be anticipated are:

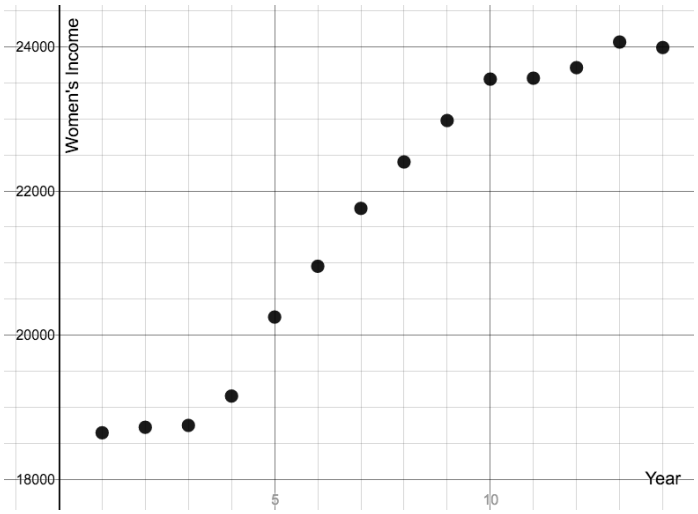
- Trying to get the greatest number of points on the line
- Selecting a point at the beginning and end of the distribution and connecting them.
- Trying to get as many points above the line as below the line.

Discuss Part One (Whole Class):

Begin the discussion by displaying the two scatter plots and briefly discussion the correlation coefficient and what it is describing about the data. The graphs are shown below.

Median Annual Income for Men from 1991-2005: r = 0.814

Median Annual Income for Women from 1991-2005 r = 0.964



Focusing on the scatter plot for men's income, ask students that have used different strategies for placing the line of best fit to share their strategies and draw their lines on the graphs. As students share different strategies ask the class to compare strengths and weaknesses of the approach in modeling the trends in the data. Ask students to compare and interpret the slope of the line of best fit that they selected.

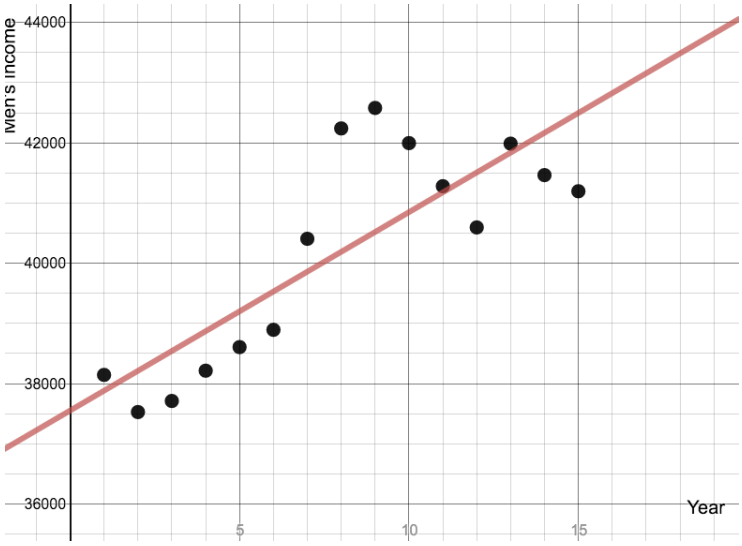
Re-launch: Demonstrate how to use technology to calculate a linear regression and graph the line along with the data on a scatter plot. Then give students time to work on the rest of the task, monitoring their discussions as they work.

Discuss (Part 2):

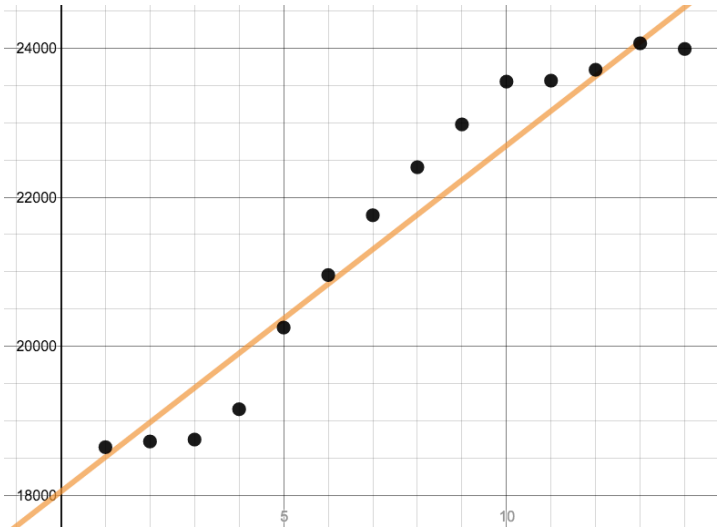
Begin the second part of the discussion by displaying the graphs and regression lines, shown below. Ask students what observations they make about the regression lines. They may be surprised that the line for the men's data doesn't actually intersect any of the data points. Ask students how they think that the regression line was calculated. Listen for students that are noticing that the lines of regression seem to cut the data "in half", leaving as much "space between the points and the line"

above and below. This will lead naturally to the idea of residuals and the way that the least squares regression line is calculated.

Regression line for men's income: $y = 329.52x + 37,554$



Regression line for women's income: $y = 464.10x + 18,053$



Ask students to consider the slopes of each of the regression lines. What does each slope mean? Since the women's slope is greater than the men's, does this mean that women make more money?

Also discuss the y-intercept on the men's income graph. The y intercept in this case is the value given by the linear model for the year 0 (1990).

Students were asked to use the model to predict the men's income for year 2015. Ask students for the predicted value based on their regression line.

After displaying the graphs that show the additional data, ask how they would modify that prediction now that they have more data. They will notice that the trend for men's income from 2005 to 2011 is downward. This highlights the danger of extrapolating, which means to extend the model well beyond the actual data.

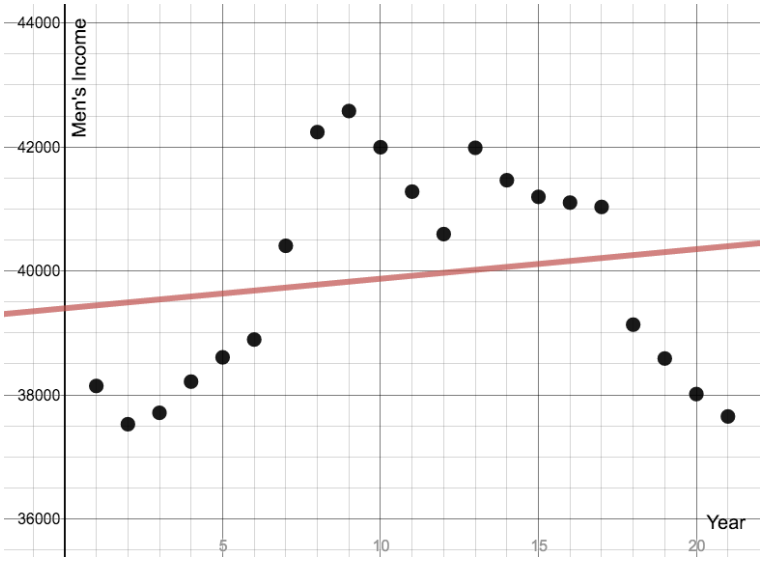
Finally, discuss the last question. Students' justification of their answer should include the use of the correlation coefficient, and compare trends in the data that can be observed in the scatter plot versus the slope of the regression line.

Graphs and linear regressions are shown below.

Median Income for Men 1991 - 2011

r = 0.17

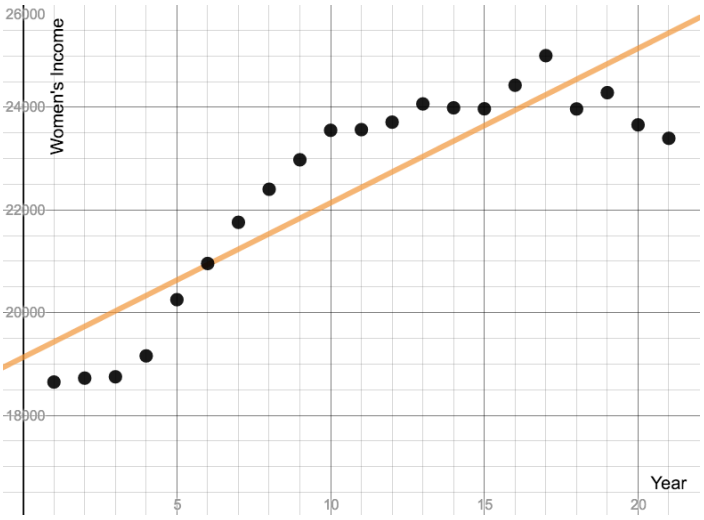
y = 47.16x + 39,398



Median Annual Income for Women 1991-2011

r = 0.88

y = 300.76x + 19131



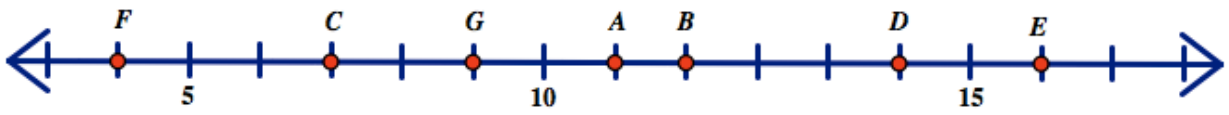
Aligned Ready, Set, Go: ***Modeling with Data 9.6***

READY, SET, GO! **Name** **Period** **Date**

READY

Topic: Finding distance and averages

Use the number line below to answer the questions.



1. Find the distance between *point A* and each of the points on the number line.

AF = _____ AC = _____ AG = _____ AB = _____ AD = _____ AE = _____

2. What is the total of all the distances from *point A* that you found in exercise number one?

3. Find the average of the distances that you found in exercise 1.

4. Which point or points on the number line is located the average distance away from *point A*?

5. Circle the location or locations on the number line that is the average distance away from *A.*

6. Find the distance between *point D* and each of the points on the number line.

DF = _____ DC = _____ DG = _____ DA = _____ DB = _____ DE = _____

7. What is the total of all the distances from *point D* that you found in exercise number six?

8. Find the average of the distances that you found in exercise 6.

9. Is there a point on the number line located the average distance away from *point D*?

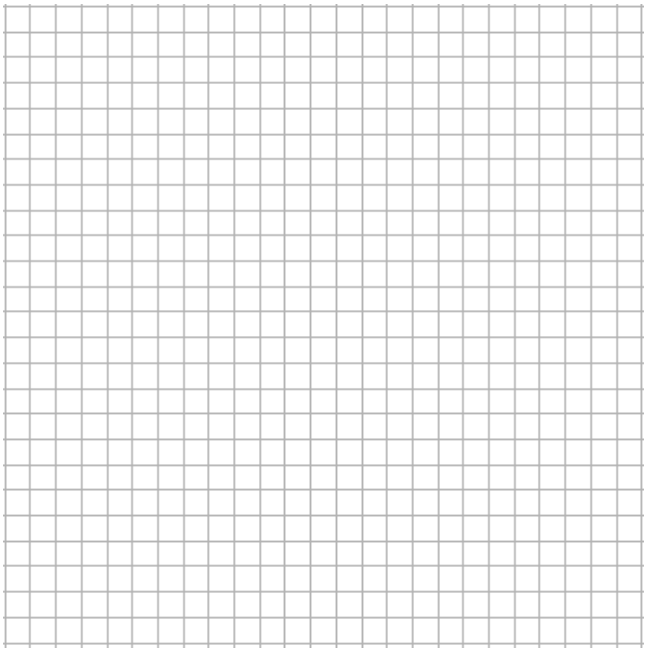
10. Label a location on the number line that is the average distance away from *point D,* label it *Y*.

SET

Topic: Scatter plots and lines of best fit or trend lines

11. Create a scatter plot for the data in the table.

English Score	History Score
60	65
53	59
44	57
61	61
70	67

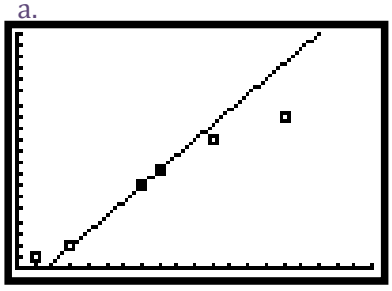


12. Do the English and history scores have a positive or negative correlation?

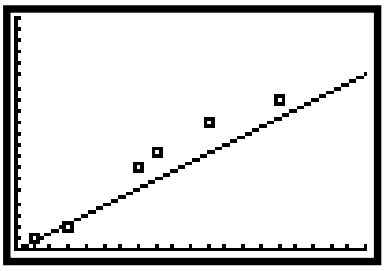
13. Do the English and history scores have a strong or weak correlation?

14. Which graph below shows the best model for the data and will create the best prediction? Explain why your choice is the best model for the data.

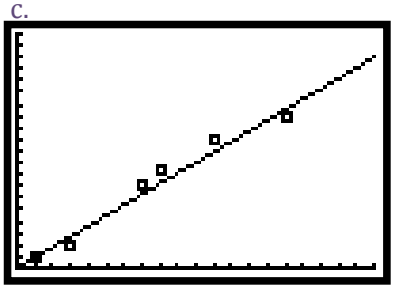
a.



b.

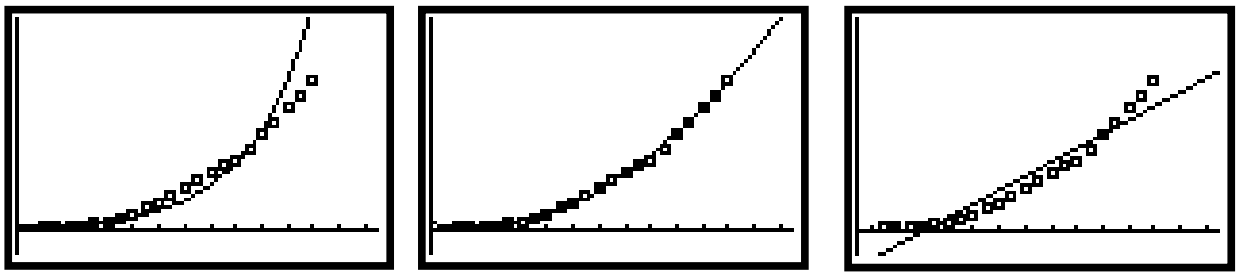


c.



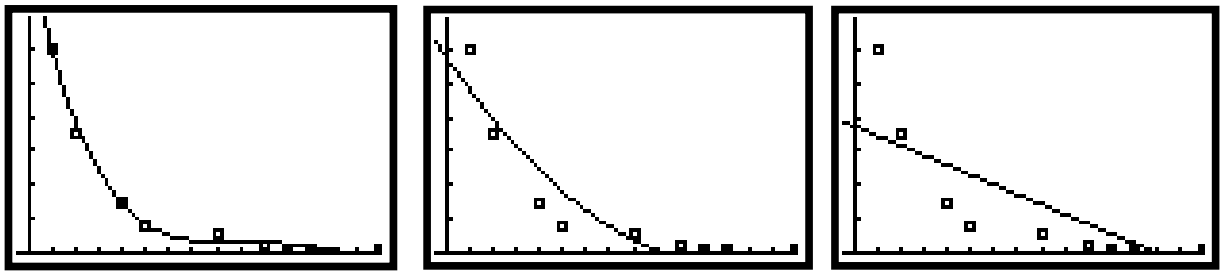
15. Which graph below shows the best model for the data and will create the best prediction? Explain why your choice is the best model for the data.

a. b. c.



16. Which graph below shows the best model for the data and will create the best prediction? Explain why your choice is the best model for the data.

a. b. c.



GO

Topic: Creating explicit function rules for arithmetic and geometric sequences.

Use the given information below to create an explicit function rule for each sequence.

17. $f(2) = 7$; common difference = 3

18. $g(1) = 8$; common ratio = 2

19. $h(6) = 3$; common ratio = -3

20. $r(5) = -3$; common difference = 7

21. $g(7) = 1$; common difference = -9

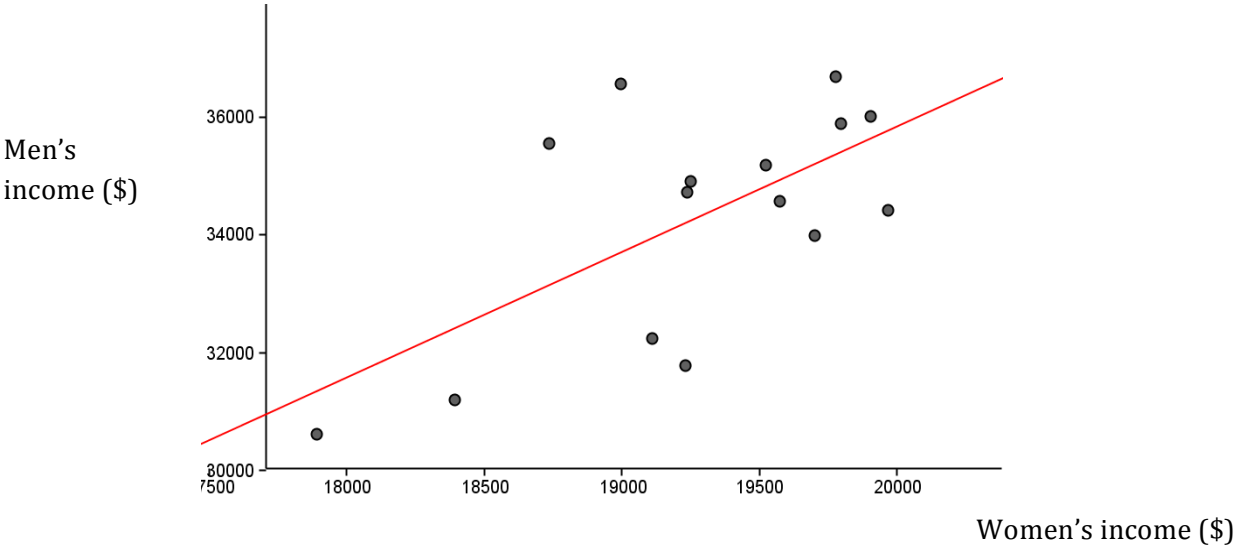
22. $g(1) = 5$; common ratio = $\frac{1}{2}$

9.7 Getting Schooled

A Solidify Understanding Task

In *Getting More $,* Leo and Araceli noticed a difference in men's and women's salaries. Araceli thought that it was unfair that women were paid less than men. Leo thought that there must be some good reason for the discrepancy, so they decided to dig deeper into the Census Bureau's income data to see if they could understand more about these differences.

First, they decided to compare the income of men and women that graduated from high school (or equivalent), but did not pursue further schooling. They created the scatter plot below, with the x value of a point representing the average woman's salary for some year and the y value representing the average man's salary for the same year. For instance, the year 2011 is represented on the graph by the point (17887, 30616). You can find this point on the graph in the bottom left corner.



1. Based upon the graph, estimate the correlation coefficient.

2. Estimate the average income for men in this time period. Describe how you used the graph to find it.

3. What is the average income for women in this time period? Describe how you used the graph to find it.

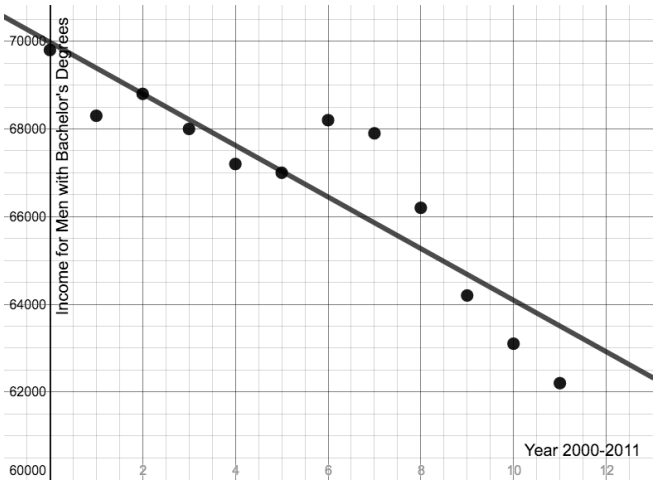
4. Leo and Araceli calculated the linear regression for these data to be $y = 2.189x - 6731.8$. What does the slope of this regression line mean about the income of men compared to women? Use precise units and language.

"Hmmmm," said Araceli, "It's just as I suspected. The whole system is unfair to women." "No, wait," said Leo, "Let's look at incomes for men and women with bachelor's degrees or more. Maybe it has something to do with levels of education."

5. Leo and Araceli started with the data for men with bachelor's degrees or more. They found the correlation coefficient for the average salary vs year from 2000-2011 was r = -.894.

   Predict what the graph might look like and draw it here. Be sure to scale and label the axes and put 12 points on your graph.

The actual scatter plot for salaries for men with bachelor's degrees from 2000-2011 is below. How did you do?



6. Both Leo and Araceli were surprised at this graph. They calculated the regression line and got $y = -588.46x + 69978$. What does this equation say about the income of men with bachelor's degrees from 2000-2011? Use both the slope and the y-intercept of the line of regression in your answer.

Next, they turned their attention to the data for women with bachelor's degrees or more from 2000-2011. Here's the data:

Year	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
Income for Women ($)	41338	42409	42746	42620	44161	44007	42690	42539	42954	42871	42992	43293

7. Analyze the data for women with bachelor's degrees by creating a scatter plot, interpreting the correlation coefficient and the regression line. For consistency with the men's graph above, use 0 for the year 2000, 1 for the year 2001, etc. Draw the graph and report the results of your analysis below:

8. Now that you have analyzed the results for women, compare the results for men and women with bachelor's degrees and more over the period from 2000-2011.

9. Leo believes that the difference in income between men and women may be explained by differences in education, but Araceli believes there must be other factors such as discrimination. Based on the data in this task and *Getting More $,* make a convincing case to support either Leo or Araceli.

10. What other data would be useful in making your case? Explain what to look for and why.

9.7 Getting Schooled – Teacher Notes

A Solidify Understanding Task

Special Note to Teachers: This task requires the use of technology that can calculate the correlation coefficient, r, and a linear regression. Most graphing calculators will work well. GeoGebra or Desmos, both powerful, free computer apps would be very helpful and easy to use on this task.

Purpose: The purpose of this task is to solidify students understanding of linear models for data by interpreting the slopes and intercepts of regression lines with various units. Students are asked to use linear models to compare and analyze data. In the task they draw conclusions and justify arguments about data. In addition they are asked to consider additional data that could be collected to inform their conclusions.

Core Standards Focus:

S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

- **a.** Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.
- **c.** Fit a linear function for a scatter plot that suggests a linear association.

S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

Standards for Mathematical Practice of Focus in the Task

SMP 3 - Construct viable arguments and critique the reasoning of others.

SMP 4 – Model with mathematics.

Launch (Whole Class):

Remind students of their work with men's and women's median annual incomes from the previous task. Ask them to recall some of the conclusions that could be made from the data. Introduce this task by telling them that they will be drawing upon their experience with correlation coefficients and linear regressions to analyze and compare data. By the time they have finished the task they should be prepared to use the data to make an argument about the differences in men's and women's salary, based upon education and other possible factors.

Explore (Small Group):

Monitor students as they work, ensuring that they are estimating as requested in the task before making the calculations. This will help to draw them into the data so that they can make sense of it and deepen their understanding. Keep students focused on using the units of slope based on the graphs. They may be more familiar with graphs that have time across the x-axis, but struggle to interpret the first graph that compares salaries of men and women where the year the data was obtained is not evident.

Discuss (Whole Class):

Actual correlation coefficient for #1 is r = 0.6421.

Begin the discussion with the meaning of the slope of the linear regression in the first graph. Students should be able to articulate the idea that the slope in this case means that the median salary for men was 2.189 times the median salary for women of the same education level. In this case the slope is a ratio of men's salaries to women's salaries or the rate that men's salaries change in relation to women's salaries.

The next slope to interpret is in #6. Students should be able to articulate that the median salary for men went down by about $588.49 each year during the time period. In this case the slope is the rate of change of men's salaries each year.

The bulk of the discussion should be an opportunity for students to dig deeply in the analysis of the data to make the case that education explains the differences in median incomes between men and

women or that there are other factors that explain the differences. Organize the class so that students are assigned to one side of the argument or the other and then take turns presenting one piece of evidence from their analysis. Record the claims and allow the other side to refute any claim that they feel is in error.

Aligned Ready, Set, Go: ***Modeling with Data 9.7***

READY, SET, GO! **Name** **Period** **Date**

READY

Topic: Finding distances and averages

The graph below shows several points and the line $y = x$. Use the graph to answer each question.

1. The vertical distance between *point N* and the line $y = x$ on the graph is 3.
   Find all of the vertical distances between the points and the line $y = x$.

   B:

   D:

   E:

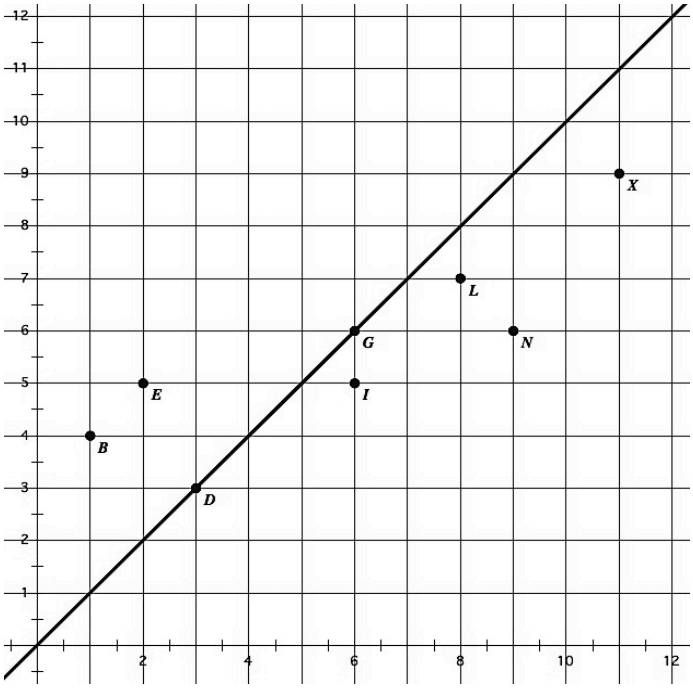
   G:

   I:

   L:

   N:

   X:



2. Calculate the *sum of all the distances* you found in exercise one.

3. What is the *average vertical distance* of the points from the line $y = x$?

4. Is the line shown on the graph the line of best fit?
   Explain why or why not.
   If it is not the best line, draw one that is better fit to the data.

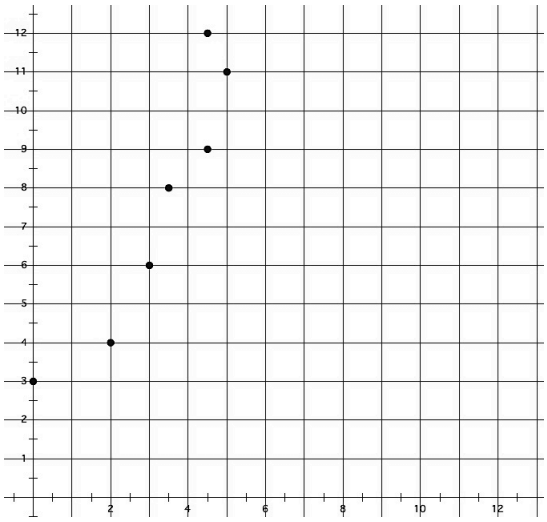
5. Estimate the correlation coefficient for this set of data points.
   If you have a way to calculate it exactly, check your estimate. (You could use a graphing calculator or data software.)

SET

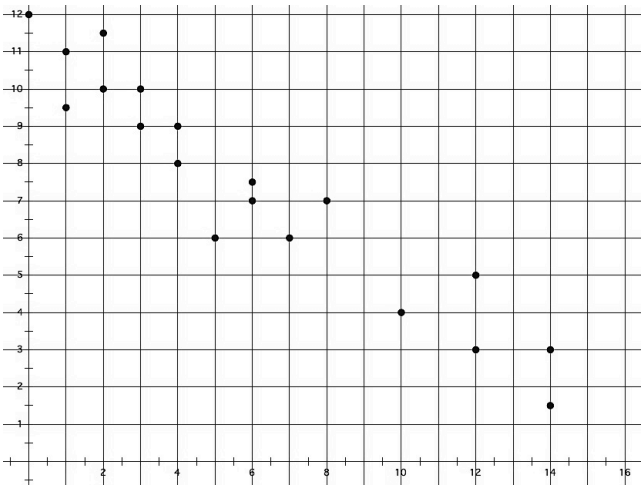
Topic: Creating and analyzing scatter plots

Determine whether a linear or an exponential model would be best for the given scatter plot. Then sketch a model on the graph that could be used to make predictions.

6.



7.



8. a) Use the data in the table below to make a scatter plot.

b) Is the correlation of the graph positive or negative? Why?

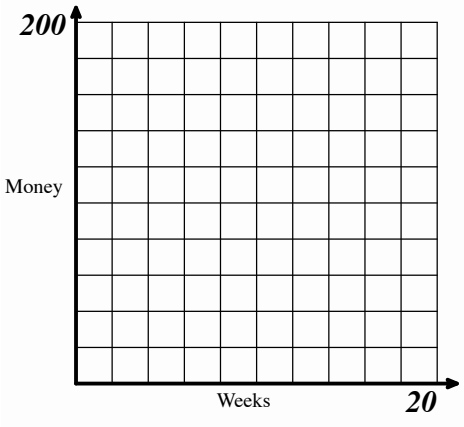
c) What would you estimate the correlation coefficient to be? Why?

d) Create a regression line and write the regression equation.

e) What does the slope of the regression equation mean in terms of the variables?

f) Most school years are 36 weeks. If the rate of spending is kept the same, how much more money needs to be saved during the summer in order for there to be money to last all 36 weeks?

Weeks since school started	Money in savings
1	200
3	175
4	162
7	120
10	87
13	57
20	5



GO

Topic: Determining when to use a two-way table and when use a scatter plot

9. In which situations does it make the most sense to use a two-way table and look at the relative frequencies.

10. In which situations does it make the most sense to use a scatter plot and a linear or exponential model to analyze and make decisions or draw conclusions?

Label each representation below as a *function or not a function.* If it is a function, label it as *linear, exponential, or neither.* If is does not represent a function, explain why.

11.

x	y
0	12
1	12
2	12
3	12
4	12

12.

x	y
1	15
2	30
3	15
2	20
1	25

13.

x	y
-6	-2
-5	-3
-4	-4
-3	-5
-2	-6

14. $y + 12x = 4$

5. $y = 3 \cdot 4^{(x-1)}$

16. The amount of medicine in the blood stream of a cat as time passes. The initial dose of medicine is 80mm and the medicine breaks down at 35% each hour.

17.

Time	0	1	2	3	4
Money in bank	$250	$337.50	$455.63	$615.09	$830.38

9.8 Rockin' the Residuals

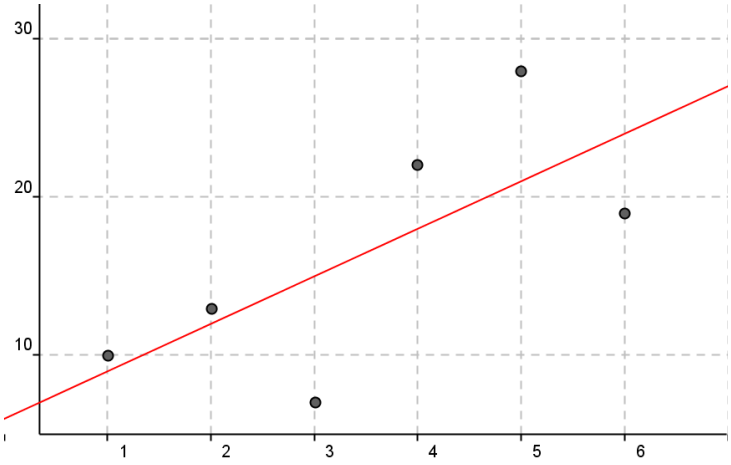
A Solidify Understanding Task

The correlation coefficient is not the only tool that statisticians use to analyze whether or not a line is a good model for the data. They also consider the residuals, which is to look at the difference between the observed value (the data) and the predicted value (the y-value on the regression line). This sounds a little complicated, but it's not really. The residuals are just a way of thinking about how far away the actual data is from the regression line.

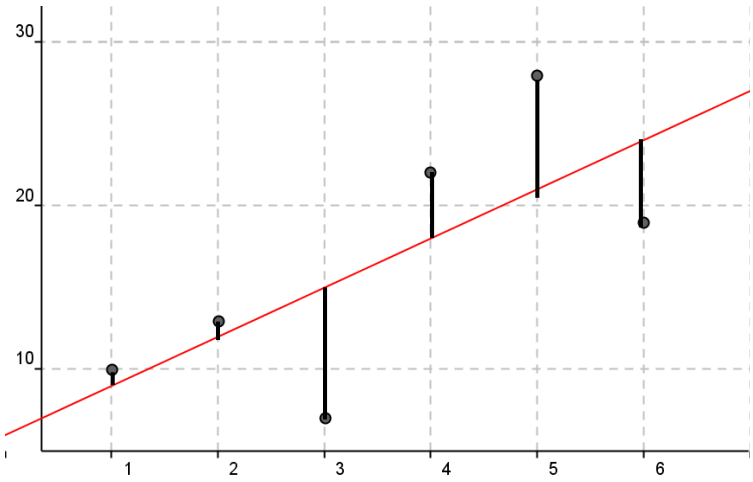
Start with some data:

x	1	2	3	4	5	6
y	10	13	7	22	28	19

Create a scatter plot and graph the regression line. In, this case the line is $y = 3x + 6$.



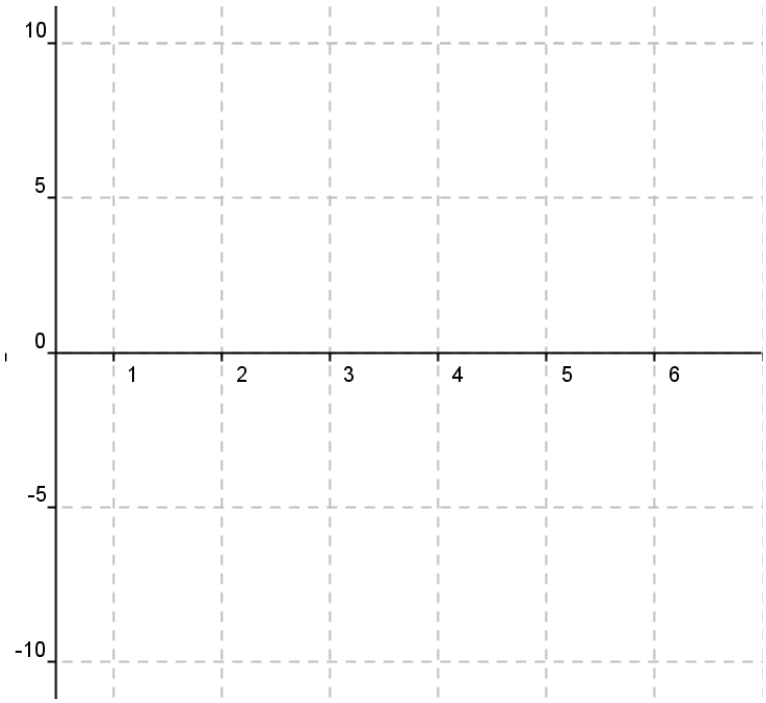
Draw a line from each point to the regression line, like the segments drawn from each point below.



1. The residuals are the lengths of the segments. How can you calculate the length of each segment to get the residuals?

2. Generally, if the data point is above the regression line the residual is positive, if the data point is below the line, the residual is negative. Knowing this, use your plan from #1 to create a table of residual values using each data point.

3. Statisticians like to look at graphs of the residuals to judge their regression lines. So, you get your chance to do it. Graph the residuals here.



Now, that you have constructed a residual plot, think about what the residuals mean and answer the following questions.

4. If a residual is large and negative, what does it mean?

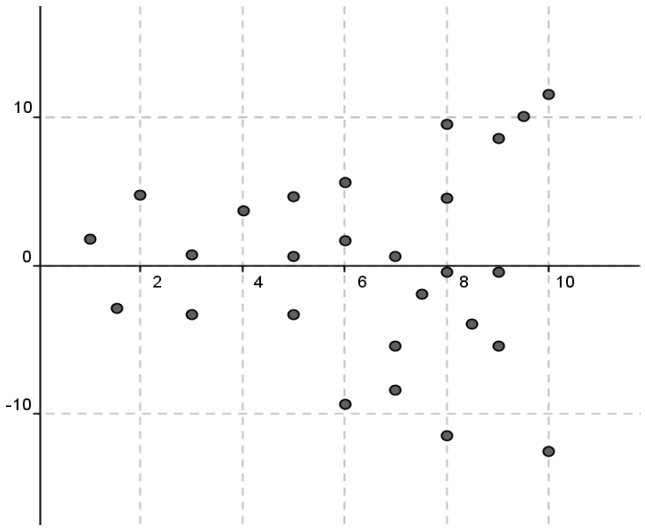
5. What does it mean if a residual is equal to 0?

6. If someone told you that they estimated a line of best fit for a set of data points and all of the residuals were positive, what would you say?

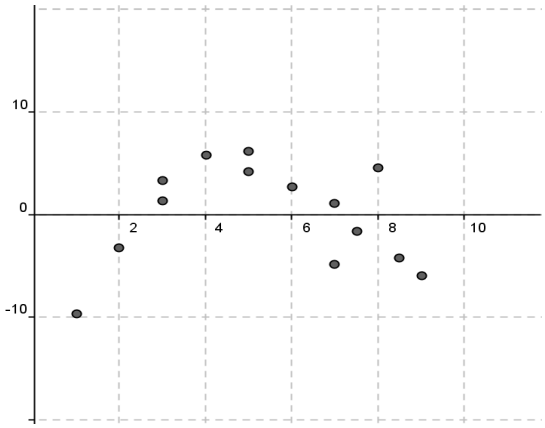
7. If the correlation coefficient for a data set is equal to 1, what will the residual plot look like?

Statisticians use residual plots to see if there are patterns in the data that are not predicted by their model. What patterns can you identify in the following residual plots that might indicate that the regression line is not a good model for the data? Based on the residual plot are there any points that may be considered outliers?

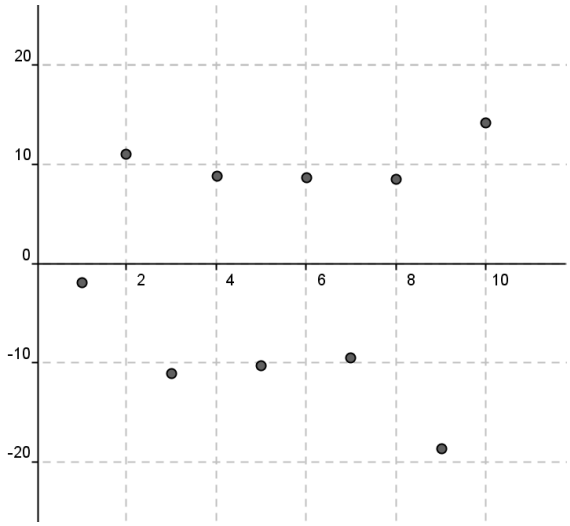
8.



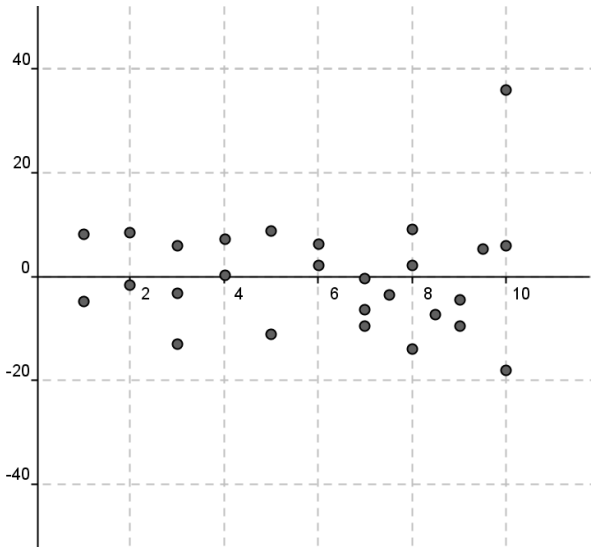
9.



10.



11.



9.8 Rockin' the Residuals – Teacher Notes

A Develop Understanding Task

Purpose: The purpose of this task is to develop an understanding of residuals and how to use residual plots to analyze the strength of a linear model for data.

Core Standards Focus:

S.ID.6: Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

S.ID.6b: Informally assess the fit of a function by plotting and analyzing residuals.

Related Standards: S.ID.6a, S.ID.6c

Standards for Mathematical Practice of Focus in the Task

SMP 7 - Look for and make use of structure.

The Teaching Cycle:

Launch (Whole Class):

Begin the task by walking through the first part of the task with students, explaining what a residual is using the graphical representation. Give students time to complete questions 1-3 and then discuss the residual plot that they have created. How does the residual plot compare with the scatter plot of the data with the regression line drawn? What information could be drawn from just looking at the residual plot if they had not seen the scatter plot and linear regression?

Explore (Small Group):

Allow students time to discuss the remaining questions and finish the task. Listen for the ways that they are making sense of the idea that the residual is the difference between the actual point and corresponding point on the linear model of the data.

Discuss (Whole Class):

Discuss each of the remaining questions. Finally, ask the class: What information is obtained from looking at a residual plot that is not given by the correlation coefficient? How do they work together to inform the analysis of bivariate quantitative data?

Aligned Ready, Set, Go: ***Modeling with Data 9.8***

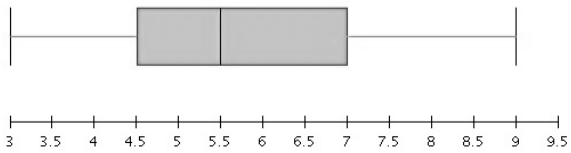
READY, SET, GO! **Name** **Period** **Date**

READY

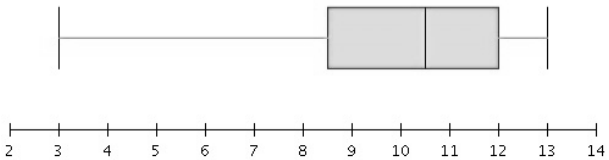
Topic: Describing spread

Describe the spread of the data set shown in each box plot shown below. Include the median, the range, and the interquartile range.

1.

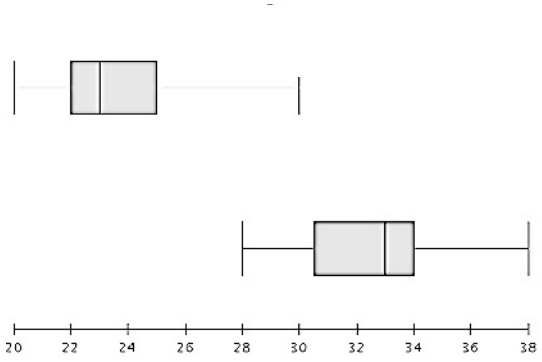


2.



3. If the box plots above represent the results of two different classes on the same assessment, which class did better? Justify your answer.

4. The two box plots below show the low temperatures for two cities in the United States. City D is the box plot on top and City E on the bottom.



a. Which city would be considered the coldest, City D or City E? Why?

b. Do these cities ever experience the same temperature? How do you know?

c. Is there a way to know the exact temperature for any given day from the box plots?

d. What advantage, if any, could a histogram of temperature data have over a box plot?

SET

Topic: Residuals, residual plots and correlation coefficients

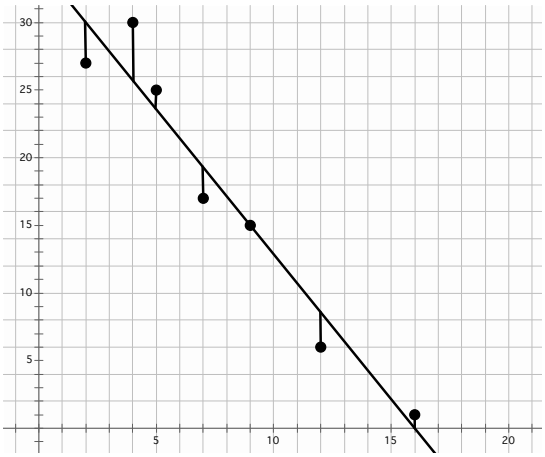
The data sheets in exercise 5 and exercise 6 are scatter plots that have the regression line and the residuals indicated. For each exercise,

a) Mark on the graph where $(\overline{x}, \overline{y})$ would be located.

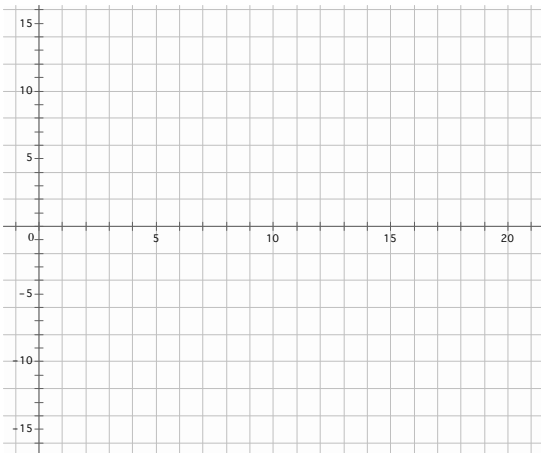
b) Use the given data sheet to create a residual plot.

c) Predict the correlation coefficient.

5. Data sheet 1 a) mark $(\overline{x}, \overline{y})$

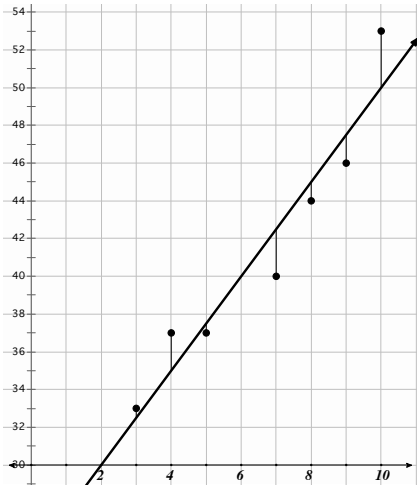


b) residual plot 1

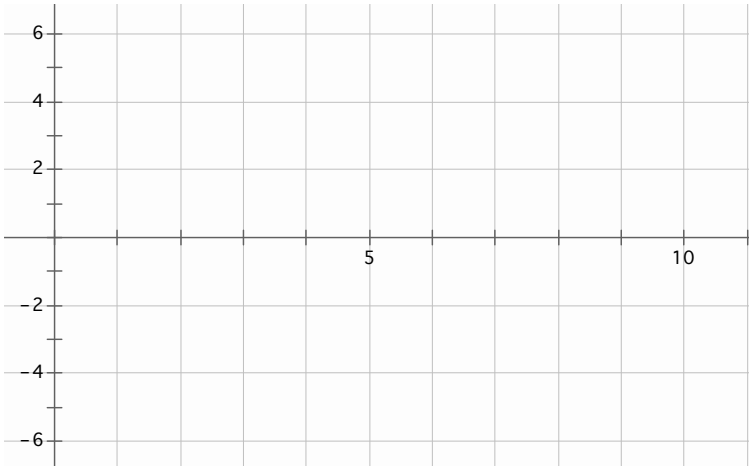


C) Correlation coefficient?

6. Data sheet 2 a) mark $(\overline{x}, \overline{y})$



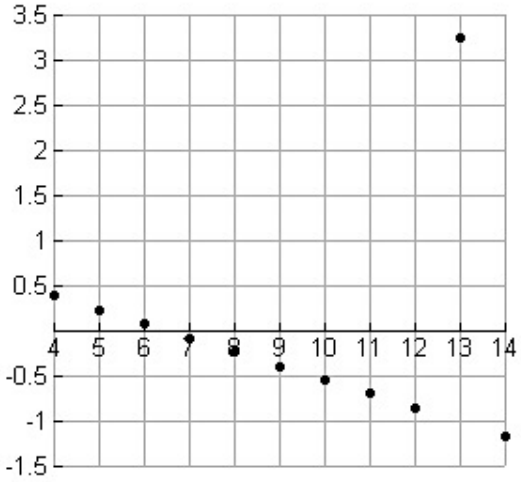
B) residual plot 2



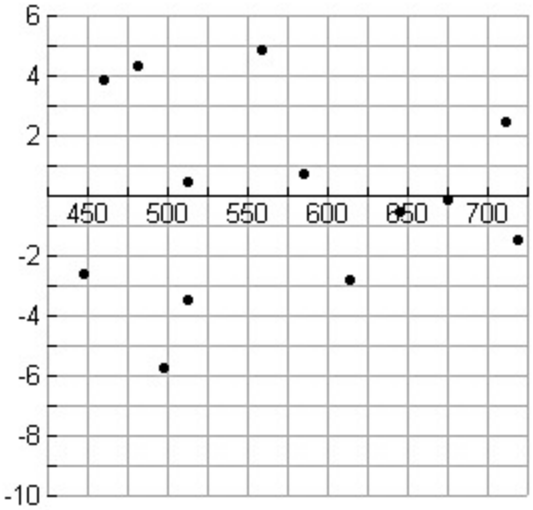
C) Correlation coefficient?

The following graphs are residual plots. Analyze the residual plots to determine how well the prediction line (line of best fit) describes the data.

7. Plot 1 analysis



8. Plot 2 analysis



GO

Topic: Geometric constructions

9. Construct an isosceles triangle with a compass and a straight edge.

10. Construct a square using a compass and a straight edge.

11. Use a compass and a straight edge to construct a hexagon inscribed in a circle.

9.9 Lies and Statistics

A Practice Understanding Task

Decide whether each statement is:

- Sometimes true
- Always true
- Never true

Give a reason for your answer.

1. The slope of the linear regression line can be calculated using any two points in the data.

   _____

2. If the correlation coefficient for a set of data is 0, then the line of best fit is horizontal.

   _____

3. The sum of the residuals for the line of best fit is 0.

   _____

4. If the correlation coefficient is very large, then there must be an outlier in the data.

   _____

5. A negative correlation coefficient means that the data points are very random and don't really fit a linear model.

   _____

6. A negative residual means that the regression line is very far from the actual data point.

   _____

7. If the correlation coefficient is positive, then the slope of the line of best fit will probably be positive.

   _____

8. If there is a perfect correlation between variables in the data, then the correlation coefficient is 1.

_____

9. To find the value of a residual for a point (a, b) given a line of best fit, *f(x):*
   a. Find $f(a)$
   b. Find $b - f(a)$
   c. If the answer is positive, then the point is above the line.
   d. If the answer is negative, then the point is below the line.

_____

10. The larger the residual for a given point, the further away the point is from the line of best fit.

_____

11. If there is a perfect correlation between two variables a and b, then either a caused b or b caused a.

_____

9.9 Lies and Statistics – Teacher Notes

A Practice Understanding Task

Purpose: The purpose of this task is to refine students' understanding of correlation coefficients, residuals, and linear regression. As students reason through the statements that have been given, they will have to consider various cases, along with considering the definition of the statistical terms used. They will make arguments to justify their answers, citing examples and definitions.

Core Standards Focus:

S-ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

a. Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.

b. Informally assess the fit of a function by plotting and analyzing residuals.

c. Fit a linear function for a scatter plot that suggests a linear association.

Standards for Mathematical Practice of Focus in the Task

SMP 6 - Attend to precision.

SMP 3 - Construct viable arguments and critique the reasoning of others.

Teaching Cycle

Launch (Whole Class): Begin by telling students that this task is an opportunity to think like statisticians and refine the way they use the terms of statistics. Their job is to test each of the statements given to determine if they are always, sometime, or never true. In every case, they need to justify their choice with examples and reasoning. Give students some time to work on their own before sharing so that they have a chance to develop their own arguments for each problem.

Explore (Small Group): As students are sharing, listen for misconceptions that may arise so that they can be shared in the class discussion. Also watch for statements that are generating disagreement, because these are opportunities for productive reasoning and engaging discussion.

Discuss (Whole Class): Begin the class discussion with any of the statements that were controversial. In every case, have students present their arguments before guiding the class to the correct answer. Problems 2, 4, and 5 often bring out misunderstandings and should be discussed. In the remaining time, work through as many other problems as possible.

Aligned Ready, Set, Go: ***Modeling Data 9.9***

READY, SET, GO! **Name** **Period** **Date**

READY

Topic: Identifying types of functions and writing the explicit equations

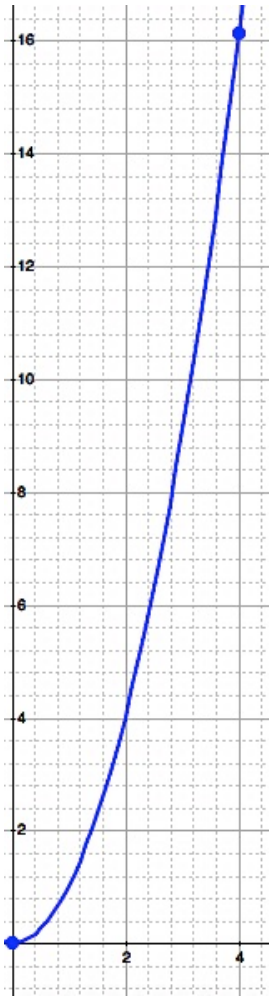
For each representation of a function, decide if the function is *linear*, *exponential*, or *neither*. Justify your answer.

1.

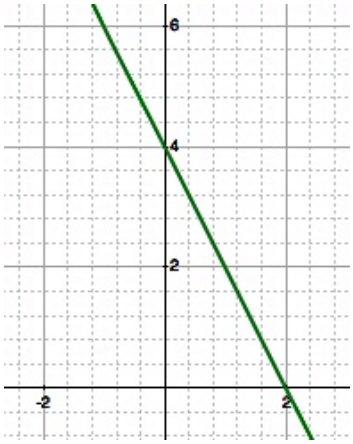
x	$f(x)$
1	117649
2	16807
3	2401
4	343
5	49

2.
The fee for a taxi ride is $7 for getting into the taxi plus $2 per mile.

3.



4.



6.

x	$f(x)$
1	1
4	2
9	3
16	4
25	5

7.
$f(1) = 7; f(x) = 5 \cdot f(x-1)$

8.
$h(x) = 3(x-1) + 2$

9.
$g(x) = 3x^2 - x - 3x^2 + 1$

SET

Topic: Reviewing key topics in statistics

Decide whether each statement is *sometimes true, always true, or never true.* If the statement is *sometimes true* give one example of when it is true and an example of when it is not.

10. The linear regression line passes through the average of the x values and the average of the y values.

11. A positive correlation coefficient means that the points in the scatterplot are very close together.

12. A negative residual means your predicted value is too low.

13. A correlation coefficient close to 1 means that a linear model is most appropriate for the data.

GO

Topic: Solving literal equations

Solve each equation for *x*.

14. $ax = d$

15. $b + cx = d$

16. $ab + cx = d$

Solve each equation for y.

17. $4x + y = 3$

18. $2y = 6x + 9$

19. $5x - 2y = 10$

Solve each equation for the indicated variable.

20. $A = \pi r^2$; Solve for r.

21. $V = \frac{lwh}{2}$; Solve for h.

22. $P = \frac{(12V)^2}{50}$; Solve for V.