Subject: Math II
Developer Name: Megan Balong
Unit Title: Correlation and Regression
Location:

Course Description:
This unit is designed for an integrated math II class but would be appropriate for any student in grades 9-12. It could be used as a supplemental unit to an existing math class or as a replacement unit for a statistical unit.

Unit Context:
This unit is designed to be a first look at correlation and a slightly more formal look at linear regression. Prior to this unit, students have experienced creating a scatter plot and finding a "best fit line" by sight, but have not used a least-squares regression line. Students should already have discussed slope and intercepts, but the unit will expand the understanding of each of these concepts. One goal of the unit is to highlight the current use of correlation and regression and the ways it may appear in daily life.

Length of Unit:
10 days and outside work

Iowa Tests Core Content Standards and Benchmarks
List all of the Iowa Tests Core Content Standards and Benchmarks that are addressed in this unit.

Core Content Standards and Benchmarks
• Students can understand and apply concepts in probability and statistics. • Students can understand and apply concepts and procedures of algebra.

ICC Essential Concepts, Skills, and Characteristics
List all of the Iowa Core Curriculum Essential Skills, Concepts, and Characteristics that are addressed in this unit.

Concepts
• Understands and applies descriptive statistics • Understands and applies inferential statistics • Understands and applies rate of change • Understands and applies functions • Reads for a variety of purposes and across content areas (literacy) • Applies writing skills and strategies to effectively communicate in a variety of genres with various audiences (literacy) • Uses writing as a tool for learning (literacy)
Skills
• Organizes and consolidates his/her mathematical thinking through communication • Uses the language of mathematics to express mathematical ideas precisely

© 2008 Iowa Department of Education

- Recognizes and applies mathematics in contexts outside of mathematics
- Uses representations to model and interpret physical, social, and mathematical phenomena

Characteristics

- Teaching for Understanding
- Problem-Based Instructional Tasks
- Mathematical Modeling
- Deep Conceptual and Procedural Knowledge
- Rigor and Relevance
- Effective Use of Technology
- Integrated Content

Big Ideas Addressed in Each Activity

Activity 1: What's Your Preference?

- Scatter plots can model the strength and direction of an association between two variables.
- The correlation coefficient can be calculated using several methods to calculate and represent the strength and direction of the association between two variables with +1 representing a perfect positive relationship, -1 representing a perfect negative relationship, and values in between representing strong, moderate, weak, or no relationship.

Activity 2: What's Your Value?

- The centroid of a set of bivariate data is always on the line of best fit and can be thought of as a representative "average" point that represents the entire set of data.
- Some types of outliers have a greater impact on the line of best fit and the strength of the correlation than others.
- Both the line of best fit and the correlation coefficient are calculated using the sum of squared errors/differences.
- A scatter plot matrix is a way to model multiple relationships between multiple variables.

Activity 3: What's the Relationship?

- The least-squares regression line is one line of best fit for a set of data.
- The least-squares regression line can be used to predict values, and its slope and y-intercept explain the relationship between the two variables.
- The accuracy of prediction based on the least-squares regression line is related to the correlation coefficient.
- The least-squares regression is found using the ideas of residuals or sum or squared errors.
- A residual plot is helpful in building a visual representation of a line of best fit.

Activity 4: Issues in Regression

- Regression is not always a straight line (linear), but also curved (non-linear).
- The correlation coefficient however, is only appropriate for linear relationships.
- The line of best fit is best used to predict values that are in between known points (interpolation). It

© 2008 Iowa Department of Education

may be appropriate to predict values based on the line of best fit that extended beyond a known point (extrapolation) for certain situations.
Activity 5: What's in the News?
o Correlation, association, and causation are not the same thing. o Correlation and regression analysis of data are used in many content areas, fields of study, and in the news. o Correlation, association, and causation are sometimes used inappropriately.
Activity 6: House Size vs. Price
o Correlation, regression, graphs, tables, and statistics can be used to judge and communicate specifically about the relationship between two variables.

© 2008 Iowa Department of Education

Rigor & Relevance Plot: *Correlation and Regression*

RIGOR

Evaluation

Synthesis

Analysis

Application

Comprehension

Knowledge

Activity 3: What's Your Relationship? Calculates and interprets residuals and least squares line. *Activity 2: What's Your Value?* Calculates and interprets the correlation coefficient QUADRANT C	*Activity 3: What's Relationship?* Calculates and interprets residuals and least squares line. *Activity 6: House Size vs Price* Analyze data for two neighborhoods for the size of home and cost. Use statistical reasoning in written analysis and comparison. QUADRANT D
Activity 4: *Issues in Regression* Discusses nonlinear regression possibilities and domain and range issues in regression analysis *Mini-summaries* during lessons and many of the practice problems throughout the unit QUADRANT A	*Activity 5: What's in the News?* Students will focus on how the word correlation is used in the newspaper. They will analyze for correct mathematical reasoning. Also, the class will discuss association vs. causation *Activity 1: What's Your Preference?* Students rank communication preferences, create a scatter plot and calculates Spearman's Rank Correlation as well as Pearson's Correlation Coefficient. Students generate understanding in a variety of contexts. QUADRANT B

Knowledge in one discipline | *Application in one discipline* | *Interdisciplinary application* | *Real-world predictable application* | *Real-world unpredictable application*

RELEVANCE

© 2008 Iowa Department of Education

Concept Map:
Include a concept map designed to provide an advanced organizer for students which shows how the big ideas build on one another and may be interconnected.

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

© 2008 Iowa Department of Education

ACTIVITY 1

Activity 1 Overview: *What's Your Preference?*

Iowa Core Curriculum Essential Skills/Concepts:	o Understands and applies descriptive statistics o Organizes and consolidates his/her mathematical thinking through communication o Uses the language of mathematics to express mathematical ideas precisely o Recognizes and applies mathematics in contexts outside of mathematics o Uses representations to model and interpret physical, social, and mathematical phenomena					
Big Ideas	o Represent bivariate data on a scatter plot. o Describe the association of bivariate data using the terminology such as strong, moderate, weak, positive, and negative. o Calculate and interpret the correlation coefficient using Spearman's Rank Correlation. o Calculate and interpret the correlation coefficient using Pearson's Rank Correlation.					
Characteristics of Instructional Core	x	Important core content that is evidence-based	x	Deep conceptual and procedural knowledge	x	Reflective teaching in a collaborative setting
	x	Teaching for understanding	x	Rigorous curriculum		Development of global perspective
	x	Teaching through problem solving and inquiry	x	Relevant curriculum	x	Incorporation of current technology
	x	Student-centered classrooms	x	Assessment for learning		
Cognitive Domain		Remembering	x	Applying		Evaluating
	x	Understanding	x	Analyzing	x	Creating
Connections to Students' Lives		Isolated within discipline		Connected to other disciplines	x	Connected to student lives
	x	Connected within discipline		Has value beyond school purposes		
Support for Literacy		Using the literacy process for inquiry		Increasing reading volume		
		Increasing access to print		Engaging students with texts		
		Involving students in discussion		Reading aloud in content areas		
		Increasing reading fluency		Explicitly instructing in vocabulary and comprehension		
		Writing to learn across content areas				
Class Time	2 class period(s) approximately 100 total minutes					

© 2008 Iowa Department of Education

Materials & Set-up:
Large graph paper, markers, large straight edge (meter stick), statistical software such as Fathom or CPMPTools, computers if using statistical software or graphing calculators, (http://www.wmich.edu/cpmp/CPMPTools/index.html), example set of rankings, example scatter plot

Teaching Tips:
This is students' first look at describing an association between two rankings. The students should develop mathematical vocabulary. Students should use large graph paper as the whole class will be using them. It is helpful to make the points on the scatter plot very dark and large as well as labeled. Once all students have completed their scatter plots have students develop a way to organize the scatter plots. Be sure that the students realize the main focus of the task is not creating the scatter plot, but how to describe and organize the scatter plots. During this class discussion it is important to use student ideas while pulling out the mathematics of correlation whenever possible. Often, students will have positive correlation. You may want to purposefully mix in examples scatter plots that have a negative correlation as well.

This activity is adapted from Core Plus Mathematics Project Course 2 Unit 4, Lesson 1, Investigation 1.

Teacher Tasks	Student Tasks
• Create an example set of rankings. • Provide motivating questions about communication preferences. • Provide an example of the scatter plot for two rankings. • Provide motivating questions for quantifying an association. • Give instructions filling out rankings. • Facilitate creation of scatter plot (be sure to check accuracy). • After all plots are done, conduct group discussion to develop consensus on descriptions of strong, moderate, weak, positive, or negative. • Ask students if they were to organize and/or quantify these scatter plots, what they would do? At this time, refrain from offering too much scaffolding here. Listen carefully to student thinking. • After discussion give mini-lecture on correlation coefficient. • Have students create scatter plot using technology and have the technology calculate the correlation coefficient.	• Fill out rankings. • Create scatter plot with partner. • Hang scatter plot on wall and decide if they think their scatter plot would be described as strong, moderate, weak, positive or negative. • Do the same for all the other groups. • Participate completely in discussion about descriptions. • With partner decide on an organizational plan. • Take notes in math toolbox on correlation coefficient. • Use technology to create scatter plot and correlation, then label the scatter plots with correlation coefficient. • Analyze how the correlation relates to scatter plots.

© 2008 Iowa Department of Education

• Conduct discussion on how the correlation coefficient relates to the scatter plots. • Assign homework for completing the activity with an adult.	

Differentiation Tasks

Less-Than-Proficient

Prioritized vocabulary, skills, or concepts for supplemental or intensive instruction:

- vocabulary: correlation, residual, strong association, moderate association, weak association, positive association, negative association
- skills/concepts: predicting from a least-squares regression line, positive and negative values between -1 and +1

Supplemental instruction suggestions:

- Prepare a number of graphs with descriptors underneath using the appropriate mathematical vocabulary
- Strategic partnerships for group work
- Create scaffolded notes for correlation coefficient mini-lecture
- Review PEMDAS
- Calculating the rank correlation by hand is not a big idea for all students. Allowing some students who are not proficient with calculations to calculate the correlation using a graphing calculator or other technology makes this activity more accessible.

Highly Proficient

Supplemental instruction suggestions:

- Pretesting may suggest the need for compacting of some activities (such as Activity 1, 2, or 4) and expansion/enrichment of others (such as Activity 5 or 6).
- Alternative assignment to find a scholarly journal article that uses rank correlation and create a brief summary of its use

Intensive instructional suggestions for students who are highly proficient

- Pre-testing may show a clear need for an alternative unit using college-level statistics materials.

Assessment:

Formative Assessments:

- Ask students if they were to organize and/or quantify these scatter plots what they would do? At this time, refrain from offering too much scaffolding here. Listen carefully to student thinking.
- Conduct discussion on how the correlation coefficient relates to the scatter plots.
- Assign homework for completing the activity with an adult.

Summative Assessments

- Provide specific suggestions for evaluating students once they have completed the activity.

© 2008 Iowa Department of Education

Reflect:

Students have enjoyed this activity in the past. It is important to maintain focus on the mathematics and the development of a system for "quantifying" an association. The second time I taught the lesson I made sure to have an example scatter plot. In addition, careful consideration to the development of mathematics through the discussion was necessary. If the conversation begins with a student with a firm understanding of rank correlation, the conversation will be stifled and students will not have time to wrestle with their own understandings.

Suggested Solutions for Activity 1

Solutions to Student Handout

Solutions to the questions throughout the activity are below. I will also assess for understanding by examining the results from the assignment.

Part One

This portion is dependent on the rankings developed from the students.

Key Points Mini-Lecture Notes

Rank correlation coefficient is a numerical value that identifies for the strength and direction of relation between two sets of data. The correlation coefficient r varies between -1 and 1. Positive correlations (between 0 and 1) indicate that high values in one set of data are related to high values in another set of data. Negative correlations (between -1 to 0) indicate that high values in one set of data are related to low values in the other set of data.

General Guidelines for the strength and significance of the coefficient

The following general categories indicate a quick way of interpreting a calculated r-value:

0.0 to 0.2 Very weak to negligible correlation
0.2 to 0.4 Weak, low correlation (not very significant)
0.4 to 0.7 Moderate correlation
0.7 to 0.9 Strong, high correlation
0.9 to 1.0 Very strong correlation

- The answers to this portion are dependent on student-developed rankings.
- The value of d will be greater and in turn the value for the entire expression becomes larger. When the numerator in a fraction becomes greater, the entire fraction becomes larger. Since the formula subtracts this value from 1, the overall value will become closer to zero.
- The value of d will be greater and in turn the value for the entire expression becomes larger. When the numerator in a fraction becomes greater, the entire fraction becomes larger. Since the formula subtracts this value from 1, the overall value will become closer to zero.

All solutions are student work dependent.

© 2008 Iowa Department of Education

Student Page/Handout

What's Your Preference?

Launch

1. Rank the following ways of communicating with your friends. Use a 1 for your favorite, a 2 for your next favorite and so on. No ties! Do not share your results with anyone, yet.

	You	Partner
Face-to-face talking		
Talking on phone		
Text messaging		
email		
Facebook/MySpace		
Instant messenger/chat rooms		

2. Now, compare with your partner. On large graph paper, plot your partner's and your rankings. Make sure yours is **as big and noticeable** as possible. Hang this graph paper up so others can see it.

   a) If you had to make an educated guess, would you say your scatter plot represents a strong, moderate, or weak association? Does your scatter plot represent a negative or positive association? Be ready explain your thinking.

   b) Looking at the set of graphs, how would you organize them? Could you put them in an order? How did you come up with your system to order them? Be ready to share your ideas with everyone.

Exploration

1. There are many ways to calculate **correlation coefficients**. In this unit, you will learn about two Spearman's Rank correlation coefficient and Pearson's correlation coefficient. We will use Pearson's correlation coefficient for much of the unit. For Pearson's correlation coefficient it is necessary to assume that both variables have a normal distribution. For Spearman's rank correlation is used if the data are ordinal or ranked or if it us unreasonable to assume that the variables are normally distributed.[1] In this activity we will focus on Spearman's rank correlation coefficient. While learning the calculations, think about why you think they work. If you are trying to measure the strength of the association between two rankings (and there are no ties), you can use **Spearman's rank correlation coefficient**. It is given by the formula:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$ where n is the number of items ranked and

$\sum d^2$ represents the sum of the squared differences between the ranks.

© 2008 Iowa Department of Education

Complete the table and use the above formula to find the Spearman Rank Correlation Coefficient.

	You	Your Partner	Difference of ranks (d)	Squared differences (d^2)
face-face				
phone				
texting				
e-mail				
Facebook				
instant message				

a) Compare your values to other students in the class. Check the calculations of at least two other partnerships.
b) Interpret your scatter plot and correlation coefficient. Write a one or two sentence summary.

2. Think about the formula for rank correlation coefficient while reading the following post from *Ask Dr. Math* (*http://mathforum.org/library/drmath/view/52774.html*) from The Math Forum. Before you read the posts, consider the following questions:
   a) When the differences between you and your partner are very large, what is going to happen to the value of $6\sum d^2$? What will happen to the overall value of the rank correlation coefficient?

   b) When the differences between you and your partner are very small, what is going to happen to the value of $6\sum d^2$? What well happen to the overall value of the rank correlation coefficient?

© 2008 Iowa Department of Education

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

© 2008 Iowa Department of Education

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

c) Find the equation of the line passing through the points (0,1) and ($\frac{N(N^2-1)}{3}$,-1).

Pearson's Correlation Coefficient is another method for calculating a measure of linear association between two pairs of values (x,y). The calculation includes standard deviation. For most sets of data, you will probably use a statistical package or graphing calculator for calculations. But it is worthy to find the correlation coefficient for a small data set by hand to better understand the formula.

3. Consider the set of data carefully calculate Pearson's Correlation coefficient.

$$r = \frac{\sum(x-\bar{x})\sum(y-\bar{y})}{(n-1)S_x S_y}$$

x	y
1	4
3	8
5	11

a) To calculate the correlation coefficient, begin by finding
   i) the mean of the x-values, $\bar{x}$
   ii) the mean of the y-values, $\bar{y}$
   iii) the standard deviation of the x-values, S_x
   iv) standard deviation of S_y

© 2008 Iowa Department of Education

b) Complete the calculation of the correlation coefficient. Recall that $\sum$ is the symbol for summation.

4. Using a statistical software package or your graphing calculator create an electronic version of your scatter plot from the launch. Find the correlation coefficient and post the correlation coefficient on your large scatter plot.

5. Looking at the class scatter plots and their corresponding correlation coefficients, what do you notice about the shape of the scatter plots and the correlation coefficients?

6. Which scatter plots had a negative correlation coefficient? Which scatter plots had a positive correlation coefficient? What might these correlation coefficients mean in the context of your rankings and your partner's rankings?

7. Using statistical software or your graphing calculator to create a set of rankings that you believe would have a correlation coefficient close to 1. Be ready to share your plot and the correlation coefficient.

8. Using statistical software or your graphing calculator to create a set of rankings that you believe would have a correlation coefficient close to -1. Be ready to share your plot and the correlation coefficient.

9. Using statistical software or your graphing calculator to create a set of rankings that you believe would have a correlation coefficient close to 0. Be ready to share your plot and the correlation coefficient.

Summarize

1. Explain what the value of the correlation coefficient represents.

2. Give a quick illustration of a scatter plot with a strong positive association and strong negative association.

3. Give a quick illustration of a scatter plot with a weak association.

Check for Understanding for Activity 1: What's Your Preference?

© 2008 Iowa Department of Education

1. Rank the following types of communication with a 1 for your favorite, a 2 for your next favorite and so on. No ties! Do not share your results with anyone, yet.

	You	Partner
face-to-face talking		
talking on phone		
text messaging		
e-mail		
Facebook/MySpace		
instant messenger/chat rooms		

2. Now, compare with your parent/guardian. On the graph paper, plot your parent's ranking and your rankings. Each point on the graph corresponds to a single type of communication.

3. How would you describe your resulting scatter plot: strong positive association, moderate positive association, weak positive association, weak negative association, moderate negative association, or strong positive association?

4. Calculate the correlation coefficient using Spearman's Rank Correlation and check your calculations with statistical software or your graphing calculator.

	You	Partner	Difference of Ranks (d)	Squared Difference (d^2)
face-to-face talking				
talking on phone				
text messaging				
e-mail				
Facebook/myspace				
instant messenger/chat rooms				
			Total d^2=	

5. Use statistical software or graphing calculator to check the correlation coefficient. Then interpret your rank correlation coefficient.

© 2008 Iowa Department of Education

ACTIVITY 2

Activity 2 Overview: What's Your Value?

Iowa Core Curriculum Essential Skills/Concepts:	o Understands and applies descriptive statistics o Organizes and consolidates his/her mathematical thinking through communication o Uses the language of mathematics to express mathematical ideas precisely o Recognizes and applies mathematics in contexts outside of mathematics o Uses representations to model and interpret physical, social, and mathematical phenomena o Uses writing as a tool for learning (literacy) o Applies writing skills and strategies to effectively communicate in a variety of genres with various audiences (literacy)					
Big Ideas	o Represent bivariate data on a scatter plot. o Describe the association of bivariate data using the terminology such as strong, moderate, weak, positive, and negative. o Calculate and interpret the correlation coefficient using Pearson method. o Calculate and interpret centroid of bivariate data. o Interpret and analyze a scatter plot matrix. o Effects of outliers on data setList the "big ideas" covered in this unit.					
Characteristics of Instructional Core	x	Important core content that is evidence-based	x	Deep conceptual and procedural knowledge		Reflective teaching in a collaborative setting
	x	Teaching for understanding	x	Rigorous curriculum		Development of global perspective
	x	Teaching through problem solving and inquiry	x	Relevant curriculum	x	Incorporation of current technology
	x	Student-centered classrooms	x	Assessment for learning		
Cognitive Domain		Remembering	x	Applying		Evaluating
	x	Understanding	x	Analyzing		Creating
Connections to Students' Lives		Isolated within discipline		Connected to other disciplines	x	Connected to student lives
	x	Connected within discipline		Has value beyond school purposes		
Support for Literacy		Using the literacy process for inquiry		Increasing reading volume		
		Increasing access to print		Engaging students with texts		
		Involving students in discussion		Reading aloud in content areas		
		Increasing reading fluency		Explicitly instructing in vocabulary and comprehension		
		Writing to learn across content areas				
Class Time	2 class period(s)		Approximately 100 total minutes			

© 2008 Iowa Department of Education

Materials & Set-up:
Statistical software such as Fathom or CPMPTools (http://www.wmich.edu/cpmp/CPMP-Tools/index.html) graphing calculators or computers, student page

Teaching Tips:
In this lesson students will solidify their visual understanding of correlation coefficient as well as build new knowledge and relationships about outliers. The activity begins with a brief visualization exercise for Pearson's Correlation Coefficient. This can be a place for the teacher to include the formula and guide students through the calculation. The alternative provided here is for students to go online and read through a conceptual explanation of the formula. Most of the time, the correlation coefficient will be calculated using statistical software or a graphing calculator. Teachers should insist on clear reasoning for each question. Since the data is generated for the discussions about outliers, teachers will want to look carefully at student work to find appropriate examples of each one. It may be helpful to have back-up data in case a student does not complete the data portion correctly.

Teacher Tasks	Student Tasks
• Create an example set of scatter plots in case of technical difficulty. • While students are guessing online, monitor if students are being thoughtful about their choices. • As a class, come to consensus about clues in the scatter plot that would indicate different types of correlation. • Answer questions about the hitting problem. Listen for description of outliers. • Check over data to see that a wide variety of ages were collected. Possibly collect a set of data for students that are of the same age. • Ask probing questions about scatter plot matrices. • Share interesting plots with entire class that will highlight different types of associations. • Discuss variability of responses. • Facilitate a discussion for types of outliers.	• Thoughtfully predict the correlation coefficient at least five times. • Share their reasoning for the why they chose the correlation coefficient. • Practice skills of interpreting correlation coefficient with the hitting problem. • Foreshadow outliers and their effect on correlation coefficient. • Ask 20 people the questions, get a good mix of ages. • Analyze scatter plot matrix. • Form questions for what they do not understand. • Analyze the plots for outliers and describe the outliers. Be ready to share descriptions with others. • Create a note sheet for scatter plot matrix and types of outliers.
Differentiation Tasks	
<u>Less-Than-Proficient</u> Prioritized vocabulary, skills, concepts: correlation, strong, moderate, weak, positive, negative, centroid	

© 2008 Iowa Department of Education

Suggestions for supplemental instruction

- Create scaffolding for how to read scatter plot matrix, teach students to trace the axes of the scatter plot matrix to its label.
- Create an outline and color-coded model for types of outliers. Demonstrate strategy of looking for outliers by 'dropping' all points along the x-axis and looking for an outlier for the independent variable and 'sliding' all points along the y-axis and looking for an outlier for the dependent variable.

Highly Proficient

Suggestions for supplemental instruction

- Have students calculate by hand the correlation coefficient and explain the purpose of the different parts of the equation
- Preview other types of correlation graphs beyond the scatter plot. Have students think about how these are useful.
- Do a quick Google search for scatter plot matrix and have students interpret color-scaled plots.

Assessment:

Formative Assessments:

- While students are guessing online, monitor if students are being thoughtful about their choices.
- As a class, come to consensus about clues in the scatter plot that would indicate different types of correlation.
- Answer questions about the hitting problem. Listen for description of outliers.
- Ask probing questions about scatter plot matrices.
- Facilitate a discussion for types of outliers.

Summative Assessments:

- Provide specific suggestions for evaluating students once they have completed the activity.

Reflect:

Students may guess at number one instead of thinking carefully about it. I needed to reinforce that the explanation of their thinking about how to match the graph to the correlation coefficient is what was important. It is a good idea to assign the data collection for number 3 as homework before you begin the activity. This way a student will be able to ask a wide range of ages. I also found it necessary to create additional examples of the scatter plot matrix. It is helpful to include guidance either by the teacher or librarian on search techniques to find the example from a book, journal, magazine or online.

Attachments:

© 2008 Iowa Department of Education

Activity 2 Student Page/Handout

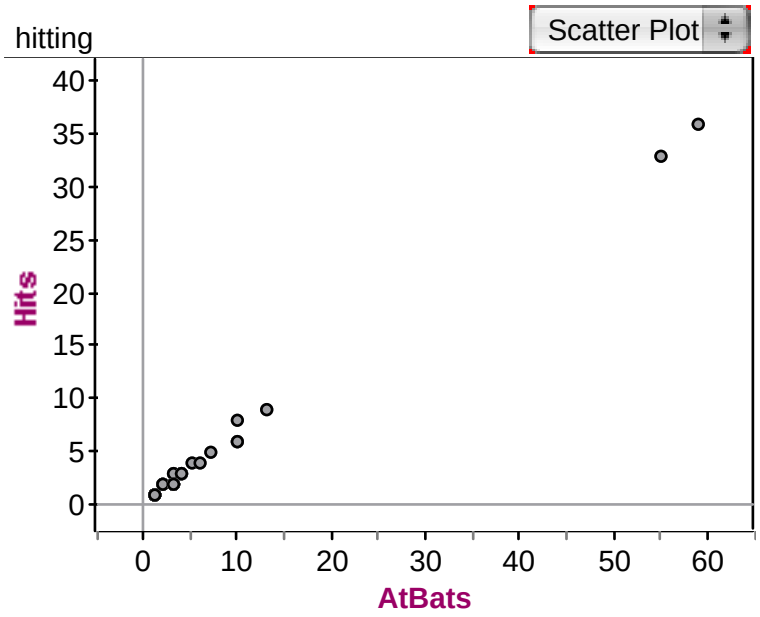
What's Your Value?

Launch

Pearson's Correlation Coefficient is a method for calculating a measure of linear association between two pairs of values (x,y). In the last activity, you calculated the correlation coefficient for a small set of data and looked at the ranges of values. Now, let's try your hand at guessing the correlation coefficient of a set of data. Go to http://www.stat.uiuc.edu/courses/stat100/java/GCApplet/GCAppletFrame.html and try at least 5 different data sets. Explain your thinking about how to match a graph and the correlation coefficient.

Exploration

1. The Iowa Girl's High school Athletic Union keeps statistics for all types of school-sponsored sports. Below are the number of "At Bats" and "Hits" for the females with the fifty best batting averages.



a. Why are there not fifty points shown?

b. Write a brief description of the scatter plot using statistical terms throughout this activity or other activities.

© 2008 Iowa Department of Education

c. What player would you want one your team? Why? Will this player have the best batting average?

d. Are there any outliers? If so, how would you describe them in relation to the rest of the data?

e. How do you think your correlation coefficient would change if you removed your outlier(s)? Predict then, try it and find out.

f. These are the leaders in the state of Iowa. How might the scatter plot change for a single team? If you can get actual statistics, create a scatter plot and compare.

2. With a partner ask at least 20 people of varying ages the following questions.
   a) What is your age?
   b) How much time per week do you typically spend using Facebook/MySpace (or any other online social network)?
   c) How much time per week do you typically spend talking or texting on your phone (cell or home)?
   d) How much time per week do you typically spending reading, composing, and sending email?
   e) How much time per week do you typically spend reading outside of work or school?
   f) How much time per week do you typically spend exercising?
   g) Create one appropriate question that has a numerical response of your choice.

© 2008 Iowa Department of Education

You will need to put the data into a spreadsheet or graphing calculator. This data will be used later in the unit, so be sure to save it.

Person	age	Facebook /MySpace usage	Phone usage (cell or home)	Time spent emailing	Time spent reading	Time spent exercising	Your question

a) Create several scatter plots using different variables you wish to be plotted. Print a copy of your scatter plots.

b) What information did you get from your scatter plots? Analyze each for associations. Write a brief description of what you notice?

c) Which variables have the strongest positive association? Which variables have the strongest negative association? Which variables have a weak association?

d) If you were to decide to look more closely at the relationship between two variables based of your scatter plots which two would you choose? Explain your reasoning.

© 2008 Iowa Department of Education

e) Do any of the scatter plots have any outliers? If so, identify what kind of outlier it is. How would you describe the outlier? Be ready to share your outlier and how you would describe it to the entire class. Together we will create a list of types of outliers.

3. Go to the following NCTM Illuminations webpage regarding methods for regression analysis and complete the questions. http://standards.nctm.org/document/eexamples/chap7/7.4/index.htm

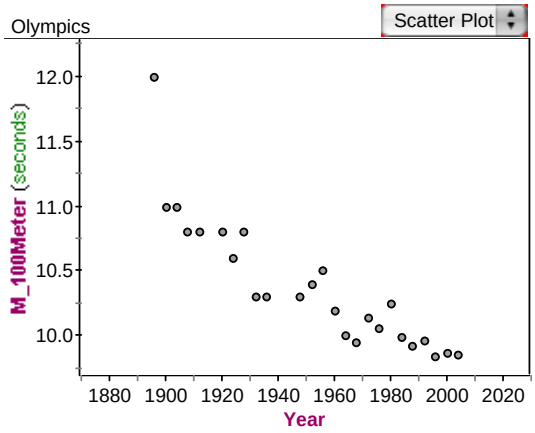
Summarize

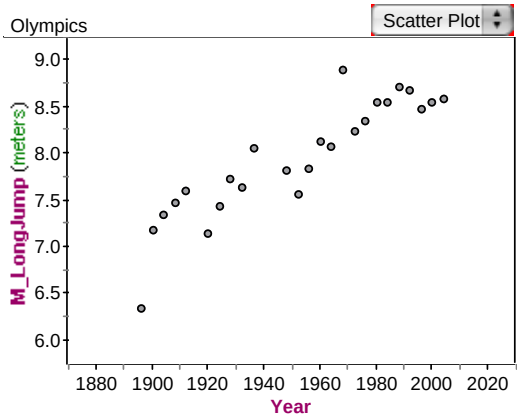
a) Explain, in your own words, why you needed to create multiple scatter plots to analyze the data.

b) What are the different types of outliers? Give an example of each.

Check for Understanding for Activity 2

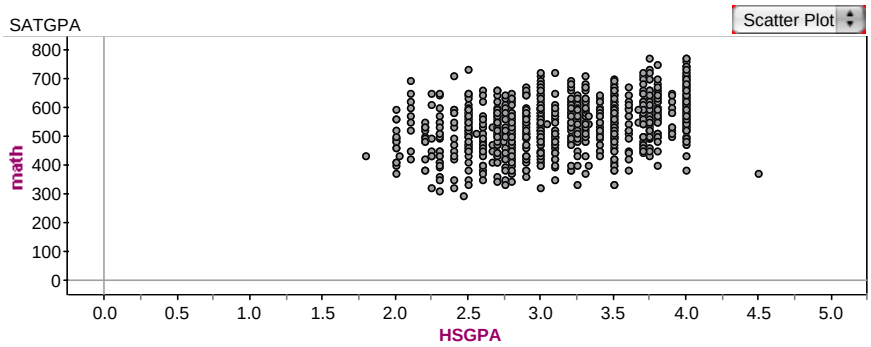
Below is a scatter plot of Olympic times for the Men's 100 meter dash and Men's Long Jump





© 2008 Iowa Department of Education

Below is a scatter plot of high school GPA and SAT Math scores.



1. Identify any outliers and describe why you consider them outliers in terms of the situation.

2. Find another example of a scatter plot in journal, magazine, and internet. Identify any outliers and interpret them in terms of the situation.

© 2008 Iowa Department of Education

Suggested Solutions to Activity 2

Possible solutions to the activity questions are given here. The formative assessment will take place through in class discussion, teacher observation and questioning. The assignment is assessed for student understanding through their written explanations and informal sharing.

1. Website is self-checking
2.
    a) Some points represent more than one data point (more than one player)
    b) There is a strong positive association between at bats and hits for the top fifty hitters in Iowa Girl's High School. The correlation coefficient is 0.99. Most of these players have not had a many times at bat probably got a hit each time, making their average close to 1.000.
    c) Most likely choices are the players to the far right. These players have had more at bats and have been successful at getting a good number of hits. Their hits to at bat ratio is high even though they have batted frequently. They probably do not have the best batting averages, because other players have been at bat fewer times and got hits each of those times making their batting average 1.00.
    d) Yes, students may consider the two points in the top right as outliers. They are outliers for the number of at bats and the number of hits.
    e) Since the correlation originally was so high, it remains very high.
    f) If you considered an entire team, the team would most likely have varying abilities at bat and would have less of a correlation between at bats and hits. The correlation coefficient would likely be weak or moderate.

3. This problem will depend on the data generated by students. There are four types of outliers to look for:
    a) An outlier for x only
    b) An outlier for y only
    c) An outlier for x and y
    d) Not an outlier for x and not an outlier for y, but an outlier when x and y are considered jointly

© 2008 Iowa Department of Education

ACTIVITY 3

Activity 3 Overview: What's the Relationship?

Iowa Core Curriculum Essential Skills/Concepts:	o Understands and applies descriptive statistics o Organizes and consolidates his/her mathematical thinking through communication o Uses the language of mathematics to express mathematical ideas precisely o Recognizes and applies mathematics in contexts outside of mathematics o Uses representations to model and interpret physical, social, and mathematical phenomena o Understands and applies functions o Understands and applies rate of change					
Big Ideas	o Models data using least-squares regression line o Understands conceptually the least-squares regression line o Interprets slope and y-intercept in context o Connects the value of the correlation coefficient to least-squares regression o Calculates and interprets residuals and errors of prediction o Interprets a residual plot o Calculates and interprets sum of squared errors o Calculates and interprets centroid of bivariate data o Effective of outliers on data set					
Characteristics of Instructional Core	x	Important core content that is evidence-based	x	Deep conceptual and procedural knowledge	x	Reflective teaching in a collaborative setting
	x	Teaching for understanding	x	Rigorous curriculum		Development of global perspective
	x	Teaching through problem solving and inquiry	x	Relevant curriculum	x	Incorporation of current technology
	x	Student-centered classrooms	x	Assessment for learning		
Cognitive Domain		Remembering	x	Applying		Evaluating
	x	Understanding	x	Analyzing		Creating
Connections to Students' Lives		Isolated within discipline	x	Connected to other disciplines	x	Connected to student lives
		Connected within discipline		Has value beyond school purposes		
Support for Literacy		Using the literacy process for inquiry		Increasing reading volume		
	x	Increasing access to print	x	Engaging students with texts		
	x	Involving students in discussion		Reading aloud in content areas		
		Increasing reading fluency		Explicitly instructing in vocabulary and comprehension		
		Writing to learn across content areas				
Class Time	2 class period(s)		Approximately 100 total minutes			

© 2008 Iowa Department of Education

Materials & Set-up:
Statistical software such as Fathom or CPMPTools, computers or graphing calculator, student page, data from Activity 2. Another resource can be found at the NCTM Illuminations website: http://illuminations.nctm.org/ActivityDetail.aspx?id=146
Finally, another resource for visualizing the effects of data and outliers on correlation coefficient can be found at http://www.whfreeman.com/applet/corr.html.

Teaching Tips:
This lesson is full of mathematical ideas. The mini-summaries are important to conduct so students collect the thoughts and ideas contained in the problems. The summary time is a good time to have students create a set of notes as well. Teachers may want to have additional data sets ready for students to find and interpret the correlation coefficient as well as the least-squares regression line. The following link is a good discussion about correlation coefficient and least-squares regression: http://janda.org/c10/Lectures/topic04/L25-Modeling.htm. In addition, the following link is an e-example from the National Council of Teacher of Mathematics, *Principles and Standards* and can be used as an extension for understanding different possibilities of "line of best fit".
http://standards.nctm.org/document/eexamples/chap7/7.4/index.htm

Teacher Tasks	Student Tasks
• Inquire about use of Facebook or MySpace. Have someone share their page if available so those students without a page can visualize. • Have students work independently or in partnerships on #1-2. • Facilitate exploration with the moveable line. • Ask probing questions about what students are noticing. • Conduct a mini-summary over #1 and #2. • Have students work independently or in partnerships on #3-6. • Conduct a mini-summary of mathematics from #3-6. • Encourage students to begin on the assignment for this activity. • Provide ideas for where to find statistics.	• Offer experiences with social networking websites. • Make prediction about relationship between friendship and wall posts. • Complete #1-2. • Carefully observe the effect of the moveable line. • Participate in mini-summary by offering ideas and creating notes. • Use algebra in predictions. • Complete the assignment thoughtful with interest in the variables chosen.
Differentiation Tasks	
<u>Less-Than-Proficient</u> Prioritized vocabulary, skills, and concepts: line of best fit/least squares regression line, slope, y-intercept, residual or error of prediction.	

© 2008 Iowa Department of Education

Suggestions for supplemental instruction

- Some students may want to complete assignment at school with support.
- For assignment, scaffold by book marking several sources of music, sports, and entertainment statistics
- Strategic partnerships
- Scaffold questions for slope in context
- Scaffold notes for mini-summaries
- Create technology outline with more step-by-step instructions (support)

Highly Proficient

Suggestions for supplemental instruction

- Strategic partnerships
- Have students pick variables of their own interest (with approval) and research them

Assessment:

Formative Assessments:

- Facilitate exploration with the moveable line
- Ask probing questions about what students are noticing
- Conduct a mini-summary over #1 and #2
- Conduct a mini-summary of mathematics from #3-6
- Encourage students to begin on the assignment for this activity.

Summative Assessments:

- Provide specific suggestions for evaluating students once they have completed the activity.

Reflect:

In the process of teaching this lesson, our students discussed the difference between residual and error in prediction. Students needed more time to understand the connection between the scatter plot and residual plot. Like some other activities, I had to maintain the focus on the mathematics so students would not "play" with the technology for too long. I think in the future I will try to do more with connection to science and the data collection about radon levels.

© 2008 Iowa Department of Education

Solutions to Student Pages

Solutions to the activity questions are provided. Formative assessment is done through listening carefully to student responses and observing the work on the computers and with partners.

1.

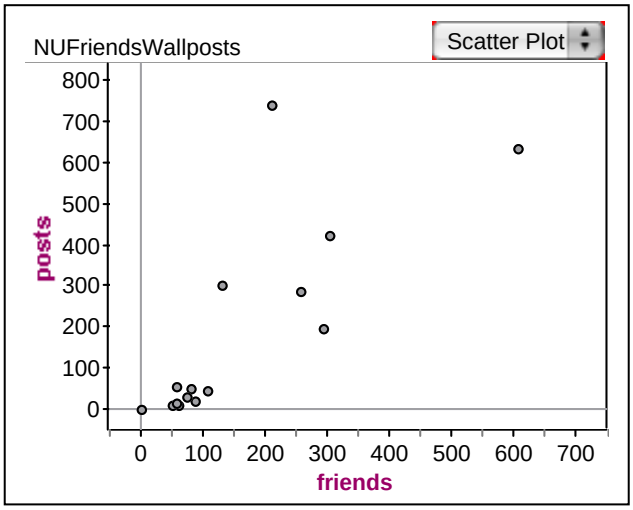
NUFriendsWallposts

	friends	posts
1	80	50
2	210	740
3	132	302
4	109	45
5	61	6
6	52	8
7	258	285
8	2	0
9	294	196
10	605	633
11	58	53
12	89	20
13	305	422
14	58	11
15	74	28

NUFriendsWallposts

	friends
posts	0.77627962

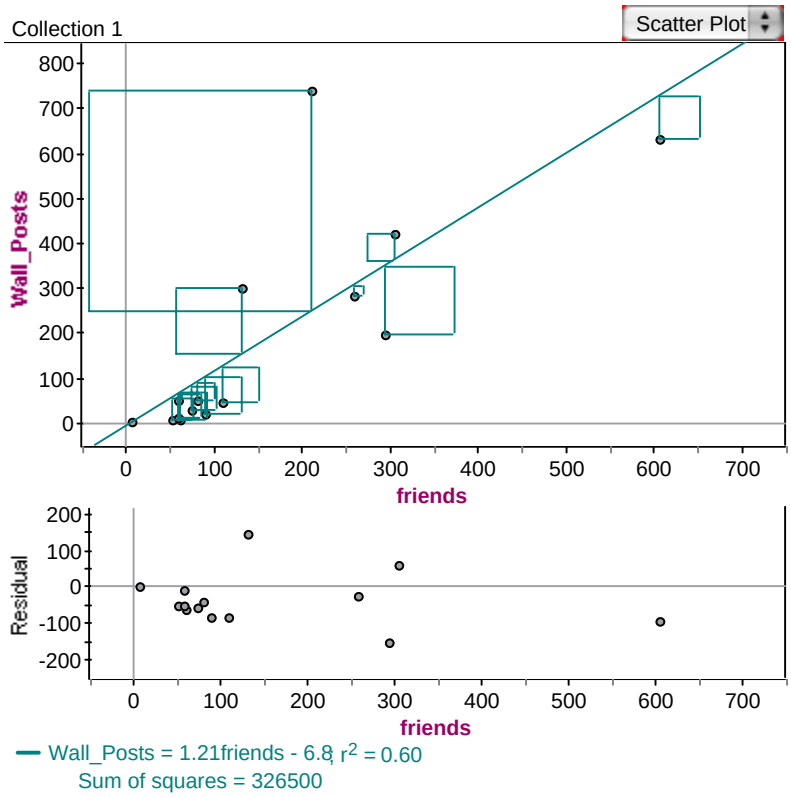
S1 = correlation ( )



4. a) The predicted value is approximates 89 wall posts, but the actual is 50 wall posts so the residual is 39 wall posts. For 210 friends, the predicted value is 247 friends, but the actual is 740, so the residual is -530 wall posts. (This is for the least-squares line. Students may have different answers because it is from the moveable line.
   b) The residual is the difference in wall posts from the predicted and the actual point.
   c) The residual plot, plots the points (friends, residual wallposts) on a scatter plot.
5. a) Slope is if a person on face book has one more friend they are likely to have 1.21 wall posts.
   b) If you had no friends, it tells how many wall posts.
   c) These do not make sense since you would not have 1.21 posts nor would you have a negative amount of friends.
   d) The predicted number of wall posts for 100 friends is approximately 116 wall posts.

© 2008 Iowa Department of Education

e) The predicted number of friends is 53.
f) No, it is to far outside the range of given data
g) The correlation coefficient is 0.77 so it is moderate to strong, but the predictions can vary greatly depending on the point. It is reasonable to make predictions, but they would have to critically look at.
h) No, this is an association not causation.
i) It is an outlier for the explanatory variable only. You could say the point (605,633) is an outlier for the x and the y variable.
j) Mean of x and mean of y (159,186.533)



6. Answers will vary. All centroids should be on the line.
7. a) Yes, because it is within the values given in the data set.
   b) No, because 200 is far away from the given data set values.
   c) It will fall on the least squares regression line.
8. a) 0.11
   b) There is a weak association between radon levels and year built.
   c) It is not useful since the correlation coefficient is so weak.

© 2008 Iowa Department of Education

What's Your Relationship?

Launch

Newsweek

Excerpt from Aug.20-27, 2007 issue

Facebook Grows Up
At 19, Mark Zuckerberg came up with a new way for college kids to connect—and started an online revolution. Now 23, he's trying to build out his business without losing its cool.

Zuckerberg himself, whose baby-faced looks at 23 would lead any bartender in America to scrutinize his driver's license carefully before serving a mojito, eschews talk about money. It's all about building the company. Speaking with NEWSWEEK between bites of a tofu snack, he is much more interested in explaining why Facebook is (1) not a social-networking site but a "utility," a tool to facilitate the information flow between users and their compatriots, family members and professional connections; (2) not just for college students, and (3) a world-changing idea of unlimited potential. Every so often he drifts back to No. 2 again, just for good measure. But the nub of his vision revolves around a concept he calls the "social graph."

As he describes it, this is a mathematical construct that maps the real-life connections between every human on the planet. Each of us is a node radiating links to the people we know. "We don't own the social graph," he says. "The social graph is this thing that exists in the world, and it always has and it always will. It's really most natural for people to communicate through it, because it's with the people around you, friends and business connections or whatever. What [Facebook] needed to do was construct as accurate of a model as possible of the way the social graph looks in the world. So once Facebook knows who you care about, you can upload a photo album and we can send it to all those people automatically."

a) Have you used an online social network? If so, what features do you use? What mathematics might be used?

b) Read the excerpt from Newsweek about Facebook. It discusses one type of mathematics Facebook uses. This activity will use data from Facebook to explore another area of mathematics.

Exploration

© 2008 Iowa Department of Education

1. One popular way to stay connected with friends is through the use of Facebook. The table below is the number of "friends" and "wall posts" for a random selection of NU students with Facebook pages.

Friends	80	210	132	109	61	52	258	2	605	58	89	305	58	74	294
Wall Posts	50	740	302	45	6	8	284	0	633	53	20	422	11	28	196

a) How would you describe the association between friends and wall posts?

b) Calculate the correlation coefficient for the set of data. What does this value indicate?

c) Are there any points that you believe effect the correlation coefficient more than others? Experiment by clicking on the point, deleting the case, and then reading the new correlation coefficient. Be sure to "UNDO" your deletion each time. Explain how you chose the points to delete and why you believe the points did or did not effect the correlation coefficient.

d) Use statistical software or a graphing calculator to create a scatter plot. Then, select graph and "Add movable line". Experiment with this line to find what you believe is the best fit. Record the equation of this line.

2. One of the ways to find the line of best fit is to calculate the least-squares regression line. In order to understand the least squares regression line, you need to first understand residuals.

Residual=observed value-predicted value.

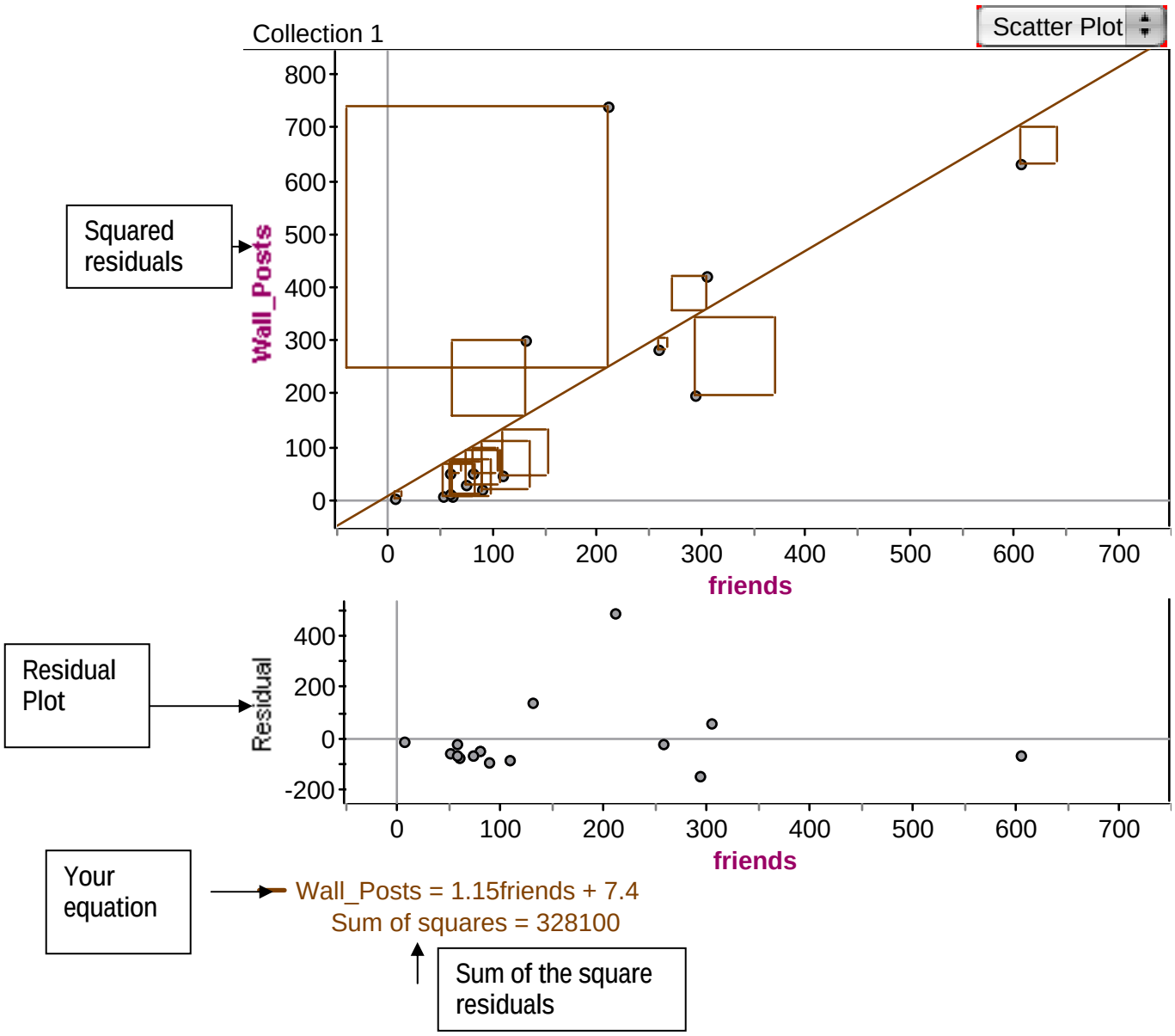
a) Find the residuals for NU Students with 80 friends and 210 friends.

b) Interpret the residuals in the context of the Facebook friends and wall posts. What do they mean?

© 2008 Iowa Department of Education

c) Graph a residual plot. What does this plot tell you?

d) Least-squares regression line is a best-fit line that is calculated by finding all the residuals, squaring the residual and summing these residuals. You can visualize this by Graph-Show Squares. Here is a screen shot of what you might see.



e) Use the moveable line to move your line to try to find where the sum of the squared residuals is the least. At the same time, be aware of the residual plot. When you believe

© 2008 Iowa Department of Education

you have found the least squares line of prediction, choose Graph-Least Squares line. What is the least squares regression line?

3. Connecting to the algebra you already know is important. Use your previous understandings to complete the following questions.
   a) What does the slope of your line mean in the context of the problem?

   b) What does the y-intercept mean in the context of your equation?

   c) Do these values make sense in the real world?

   d) You can use the least squares regression line to make predictions (dependent variable) that fall within the smallest and largest values for your explanatory variable (independent variable). Think of how many friends you have on Facebook and predict how many wall posts you might have. If you do not have a Facebook page, find the predicted number of wall posts if you have 100 friends. Show all of your work.

   e) Using symbolic reasoning, how would you use the least-squares line to estimate the number of friends given the number of wall posts that you currently have. If you do not have a Facebook page, find the number of friends a person might have if they have 58 wall posts? Show all of your work.

   f) Do you think it is reasonable to predict the number of wall posts for 1,000 friends using the least squares regression line?

© 2008 Iowa Department of Education

g) How much do you trust the predictions? How can you use the correlation coefficient to justify your answer?

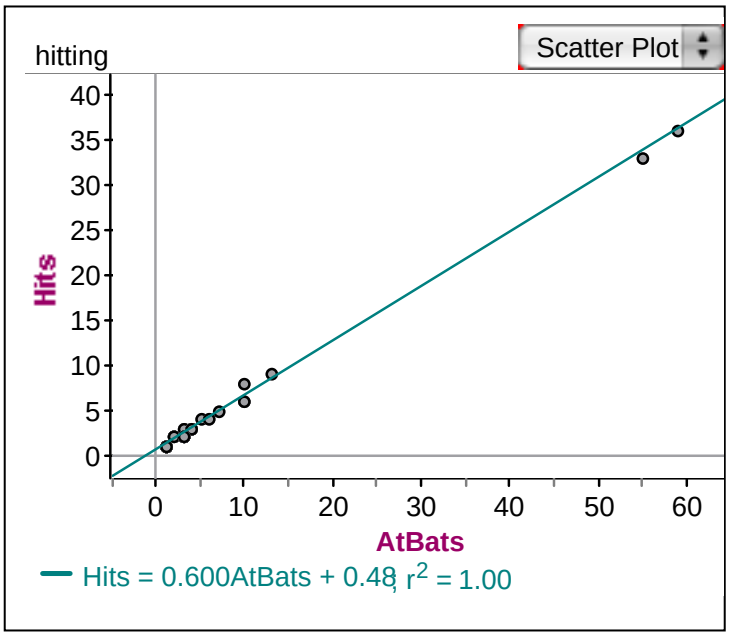
h) Does having more friends cause more wall posts?

i) Locate the point with the largest residual. Would you consider it an outlier? Why or why not? If you do consider it an outlier, how would you classify it?

j) The **centroid** is the point $(\overline{X},\overline{Y})$. Calculate this point and graph it on the scatter plot.

4. Look back at the previous activities in this unit. Find the centroid for two data sets and plot the point on the scatter plot with the least-squares regression line.

5. In the previous activity, you calculated the correlation coefficient for the at bats and hitting for Iowa girl's softball players. Here is the same scatter plot with the least-squares regression line included.



hitting

	AtBats
Hits	0.9984711

S1 = correlation ( )

a) Would it be reasonable to use the equation for the least-squares regression line to predict how many hits a player did if he/she had 30 at bats? Why or why not?

© 2008 Iowa Department of Education

b) Would it be reasonable to use the equation for the least-squares regression line to predict how many hits a player did if he/she had 200 at bats? Why or why not?

c) Given your earlier calculations with the centroid where do you predict it will be when it is plotted? Explain why you believe your prediction is true.

6. Sophomores at Northern University High School have been given the results of radon level for homes tested in the Cedar Valley. They are asked to analyze the data. The scatter plot shows the results of year the homes were built and radon levels.

   a) Out of the five correlation coefficients given as options, which is the best choice? Why?

   b) What should the sophomores conclude about the association of year homes were built and radon levels?

   c) The equation of least-squares regression line is y=0.02x-40.07. How useful is the line for predicting radon levels in Cedar Valley homes if you know the year the home was built?

© 2008 Iowa Department of Education

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

7. It is enlightening to at least once, calculate the least squares regression line equation by "hand". Use the following steps to help calculate the least-squares regression equation.

X	Y
3	7
2	3
4	8

Calculate the least square regression line using the following steps.

a) Calculate the mean, $\bar{x}$, and standard deviation of X, s_x.

© 2008 Iowa Department of Education

b) Calculate the mean, $\bar{y}$, and standard deviation of Y, s_y.

c) Calculate the correlation, *r,* of the data sets.

d) The slope of the least-squares regression line is $b = r(\frac{s_y}{s_x})$.

e) The y-intercept of the least-squares regression line is $a = \bar{y} - b\bar{x}$.

*Steps taken from leastsquareshelp.doc for Math 2830.

Summarize

1. Why is the line of best fit called the least-squares regression line?

2. How does the least-squares regression line describe the relationship between two variables?

3. How does the correlation coefficient relate to the least-squares regression line?

Check for Understanding for Activity 3

1. In the last activity you chose two variables that you would be interested in researching. Find more data about those two variables either by sampling more people are using data sets found in other resources. The Internet is full accessible statistics. For these variables complete the following:
   a) Create a scatter plot.
   b) Calculate and interpret the correlation coefficient.
   c) Add a moveable line, using the "squares" and the residual plot, predict the best-fit line. Record this equation.
   d) Find the least squares regression line and interpret its slope and y-intercept in terms of the context. Be specific about what the information tells you.
   e) Calculate the centroid and plot this point on your scatter plot.
   f) Locate the point with the largest residual. Would you consider it an outlier? Why or why not? If you do consider it an outlier, how would you classify it.

Additional Exploration

2. The following is adapted from NCTM Illuminations website: Finding the Relationship Between the Regression Line and the Centroid. Use the website http://illuminations.nctm.org/LessonDetail.aspx?ID=L457#whatif or Fathom or CPMPTools.

© 2008 Iowa Department of Education

a) Plot two points and determine the centroid of these two points.
b) Compute the midpoint of the two points. Compare this midpoint to the centroid you computed. Explain the connection.
c) Using software plot three points.
   i. Determine the centroid of these three points without technology.
   ii. Show the centroid. Confirm that you get the same point as in part (i).
   iii. Show the least-squares regression line. What happens?
d) Experiment with several scatter plots to find the relationship between the centroid and the least-squares regression line. Describe this relationship.

© 2008 Iowa Department of Education

ACTIVITY 4

Activity 4 Overview: Issues in Regression

Iowa Core Curriculum Essential Skills/Concepts:	o Understands and applies descriptive statistics o Organizes and consolidates his/her mathematical thinking through communication o Uses the language of mathematics to express mathematical ideas precisely o Uses representations to model and interpret physical, social, and mathematical phenomena o Understands and applies functions					
Big Ideas	o Understands appropriate uses of regression and correlation o Models data using different function families o Interpolation and Extrapolation					
Characteristics of Instructional Core	x	Important core content that is evidence-based	x	Deep conceptual and procedural knowledge		Reflective teaching in a collaborative setting
	x	Teaching for understanding	x	Rigorous curriculum		Development of global perspective
	x	Teaching through problem solving and inquiry		Relevant curriculum		Incorporation of current technology
	x	Student-centered classrooms	x	Assessment for learning		
Cognitive Domain		Remembering	x	Applying		Evaluating
	x	Understanding		Analyzing		Creating
Connections to Students' Lives		Isolated within discipline		Connected to other disciplines		Connected to student lives
	x	Connected within discipline		Has value beyond school purposes		
Support for Literacy		Using the literacy process for inquiry		Increasing reading volume		
		Increasing access to print		Engaging students with texts		
		Involving students in discussion		Reading aloud in content areas		
		Increasing reading fluency		Explicitly instructing in vocabulary and comprehension		
		Writing to learn across content areas				
Class Time	1 class period(s) Approximately 50 total minutes					

Materials & Set-up:
Statistical software such as Fathom or CPMP-Tools, student page, possibly M&Ms, a paper plate, and a cup for each group

Teaching Tips:
Regression can be linear or nonlinear, so this is a good opportunity to review function families and domain and range. This activity also allows for the conversation about association and causation to begin. This will be explored more in the next activity. It is critical to have students share their nonlinear examples in a meaningful way. These should generate deeper understanding about nonlinear regression. The teacher

© 2008 Iowa Department of Education

may want to collect some examples. If there is access to scientific probes, it may be a good time to collect data and find the best-fit regression equation.

Teacher Tasks	Student Tasks
• Launch discussion about regression and linear equations. Can you have a regression equation that is not linear? Can you have a correlation coefficient that does not describe a linear relationship? • Have students select 1A or 1B depending on their interests. • Ask probing questions about regression. • Give mini-summary of coefficient of determination • Have students complete #2. • Introduce #3. • Facilitate finding examples of nonlinear regression. • Supply suggestions for sources of data. • Conduct class discussions for mini-summary of nonlinear regression.	• Participate in the launch discussion. • Complete 1A or 1B. • Participate in mini-summary. • Critically think about issues of regression. • Find nonlinear regression example, association and causation. • Share examples using correct mathematical vocabulary. • Solidify understanding of correlation and regression.

Differentiation Tasks

<u>Less-Than-Proficient</u>

Prioritized vocabulary, skills, concepts: nonlinear regression, interpolation, extrapolation

Suggestions for supplemental instruction

- Have a list of function families and their parent functions to use in referring to scatter plot shapes
- May do 1A and 1B as station teaching if co-teaching
- Create a keyword picture for interpolate and extrapolate

<u>Highly Proficient</u>

Suggestions for supplemental instruction

- Investigate other relationships than exponential and/or use of logarithms in modeling for exponentials

<u>Assessment:</u>

Formative Assessments:

- Launch discussion about regression and linear equations. Can you have a regression equation that is not linear? Can you have a correlation coefficient that does not describe a linear relationship?
- Ask probing querstions about regression.
- Give mini-summary of coefficient of determination.

© 2008 Iowa Department of Education

- Facilitate finding examples of nonlinear regression
- Conduct class discussions for mini-summary of nonlinear regression

Summative Assessments:

- Provide specific suggestions for evaluating students once they have completed the activity.

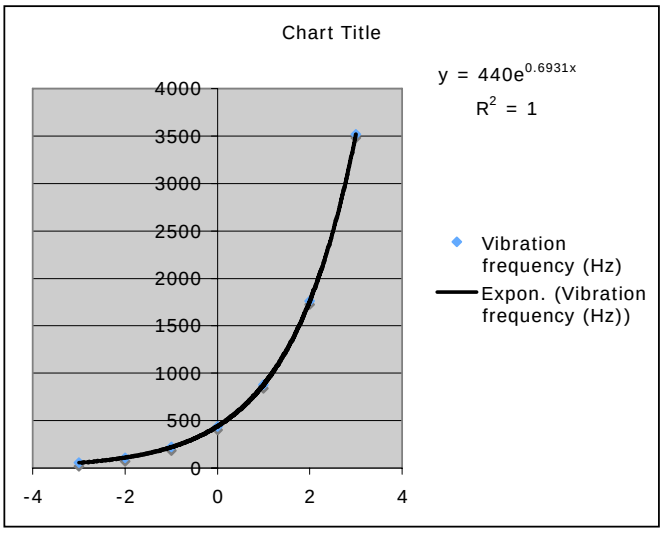
Reflect:

After teaching the unit, the coefficient of determination may not be a key topic for all students. I wanted all students to realize that the correlation coefficient is a measure of strength of a linear relationship. I will be careful in structuring the first question that students do not go to far into a statistical topic in which they may not be ready to engage. Students seem to struggle to find the reasoning for 2b. This question may need some additional teacher questioning. Again, in the future I will try to connect to science class to see what is being studied that may illuminate the concepts.

Suggested Solutions

Solutions for the activity questions are provided.

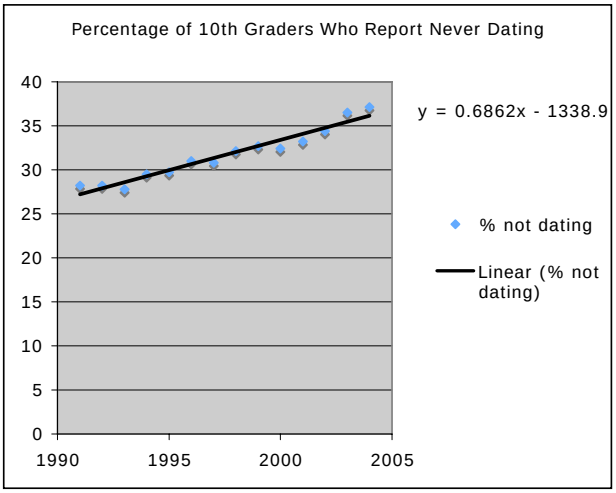
1. Both activities A and B will provide an exponential model. Here is the data for A. The data for B is student generated

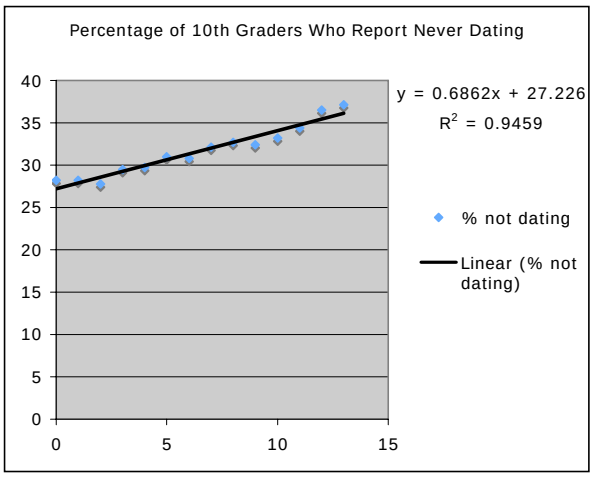


This equation can be written as $y=440*2^x$ where x = # octaves above or below A 440.

© 2008 Iowa Department of Education

2. The equation will not make sense until the domain is adjusted.

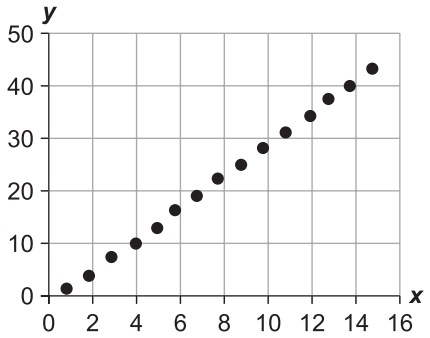




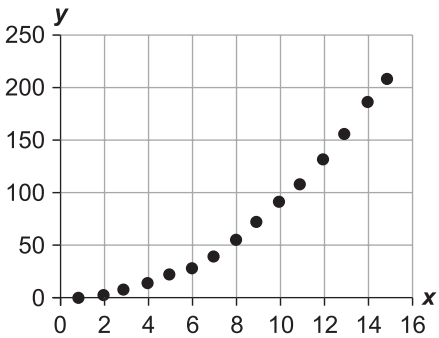
3. Students should have the following values in their tables.

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$y = 3x - 2$	1	4	7	10	13	16	19	22	25	28	31	34	37	40	43
$y = x^2 - x$	0	2	6	12	20	30	42	56	72	90	110	132	156	182	210
$y = \frac{10}{x}$	10	5	3.3	2.5	2	1.7	1.4	1.3	1.1	1	0.9	0.8	0.8	0.7	0.7

Their graphs should look like these.

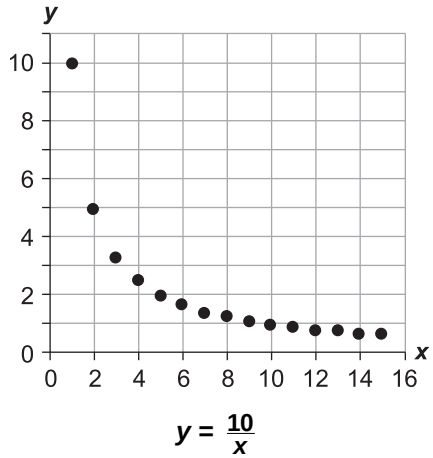


$y = 3x - 2$



$y = x^2 - x$

© 2008 Iowa Department of Education



The points on the first graph lie in a straight line that slopes up from left to right, and as expected the correlation is 1.0. In the second graph, the points slope up from left to right, lying on a parabola. Nevertheless, the correlation is a high 0.97. In the graph of $y = \frac{10}{x}$, the y values of the points decrease from left to right, suggesting a moderate, negative correlation. In fact, the correlation is –0.75.

Using the quadratic equation students should get a correlation coefficient closer to zero using more values of x which are closer to 0. Sets of ordered pairs generated by linear equations will have perfect 1.0 or –1.0 correlations. Those generated by other types of equations will not be perfectly correlated. The magnitude of the correlation will depend on how much the data vary from a straight line, which may also depend on the *domain* under consideration. For example, when considering $y = \frac{10}{x}$ you may suggest that your students drop (1, 10) from the data and compute the correlation. It will be –0.85, a greater magnitude than before. They also may drop (2, 5), and the correlation (–0.91) will be even stronger.

The correlation coefficient is a statistic that indicates how closely a set of paired data fits a straight line. Although an equation like the two nonlinear ones in this task completely determines the value of y for any value of x, the ordered pairs it generates will only be "highly correlated" to the extent that the ordered pairs come close to the least squares regression line.[2]

© 2008 Iowa Department of Education

Student Handout

Activity 4: Issues in Regression

Launch

Pearson's Correlation Coefficient is a measure of the linear association between the horizontal variable and the vertical variable. It gives information about how tightly packed the data points are about the regression line. It thereby also gives information about how well the regression line fits the data. The r-values can range from -1 (strong negative linear association) to 0 (no linear association) to +1 (strong positive linear association). But beware! You will see below that the correlation coefficient, r, is sometimes misleading. You should always look at the scatter plot and combine that knowledge with the r-value in order to draw valid conclusions about the strength of the linear association.[3]

a) Is there a way to have a strong positive correlation coefficient yet not have a good line of prediction?

Exploration

Students can choose to do either 1A or 1B.

1A. There is a relationship between musical octaves and vibration frequency. The vibration frequency of the note A above middle C is 440 cycles per second (or 440 hertz). The table below gives the number of octaves above or below A and the vibration frequency.

Number of Octaves above or below A	Vibration frequency (Hz)
-3	55
-2	110
-1	220
0	440
1	880
2	1760
3	3520

a) What is the independent variable (explanatory variable)? What is the dependent variable (response variable)? Explain.

b) Create a scatter plot of your data and find the correlation coefficient.

c) Does the least-squares regression line give the best fit?

© 2008 Iowa Department of Education

d) What function family would best model the data in the scatter plot. Using statistical software or a graphing calculator, create a model of the graph.

1B. Shake the 100 m & m's in the bowl, and then carefully pour them onto the paper plate. Count the number of m & m's that landed with the "m" face up. After you have counted and recorded the number, remove these m & m's. Repeat these steps until you have removed all of the m & m's. Each time you repeat these steps, make sure you record the trial number and the amount of m & m's that landed face up.

a) Make a table of your data

b) What is the independent variable (explanatory variable)? What is the dependent variable (response variable)? Explain.

c) Create a scatter plot of your data and find the correlation coefficient.

d) Does the least-squares regression line give the best fit?

e) What function family would best model the data in the scatter plot?

f) Using statistical software or a graphing calculator, create a model of the graph.

2. The percentage of teens that do not date at all has risen steadily since the early 1990s, reaching new highs in 2004. Among 10th graders, for example, rates rose from 28 percent in 1991 to 37 percent in 2004.

Percentage of Students in 10th grade who report they have never dated 1991-2004

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	20
% not dating	28.2	28.2	27.8	29.5	29.7	31	30.8	32.1	32.7	32.4	33.2	34.4	36.5	37

*Retrieved from Child Trends Data Bank http://www.childtrendsdatabank.org/peer.cfm

a) Is there a trend for teens not dating (10th graders)? If so, find the trend line and interpret the equation.

© 2008 Iowa Department of Education

b) Does this equation make sense? Why or why not?

c) How could you fix the problem if necessary? Show your new results.

d) **Interpolation** is the process of predicting from values within the data set and **extrapolation** is the process of predicting from values outside of the data set. Would it make sense in this situation to predict the value for the percentage of teens that reported never dating in 2005? What about 2050?

3. In previous units, you have studied the relationship between variables in different ways. This unit was about association or correlation between two variables. The purpose of this task is to explore how these different ideas of relationship or association are connected to one another. Consider the following algebraic equations.

   a. $y = 3x - 2$ b. $y = \frac{10}{x}$

   Construct a table of (x, y) pairs for each equation by letting $x = 1, 2, 3, \ldots, 15$ and finding y. Then plot the data pairs for each equation on three separate graphs. Compute the correlation coefficient for the 15 pairs in each table. For which of the equations would you expect to get a high correlation? A low correlation?

   c. $y = (x-10)^2 - 1$

   Repeat the above steps for equation (C) for $x = 11, 12, 13 \ldots 20$. Compute the correlation coefficient. Then, compute the correlation coefficient if you include x=5…20. Write a report that includes the tables, graphs, correlation coefficients, and answers to the above questions. Explain the results.[2]

© 2008 Iowa Department of Education

Summarize

1. What is the difference between interpolation and extrapolation?

2. What are the issues in regression addressed in this activity?

Check for Understanding for Activity 4

Find an example of nonlinear regression to share with the class. This can be on any appropriate topic of your choosing. Be sure to include the source of the data. Do you think that your explanatory variable causes a change in your response variable? Why or why not?

© 2008 Iowa Department of Education

ACTIVITY 5

Activity 5 Overview: *What's in the News?*

Iowa Core Curriculum Essential Skills/Concepts:	o Understands and applies descriptive statistics o Organizes and consolidates his/her mathematical thinking through communication o Uses the language of mathematics to express mathematical ideas precisely o Recognizes and applies mathematics in contexts outside of mathematics o Uses representations to model and interpret physical, social, and mathematical phenomena o Reads for a variety of purposes and across content areas (literacy)					
Big Ideas	o Distinguishes between association and causation o Critical reader of information o Able to find and use data for statistical analysis					
Characteristics of Instructional Core	x	Important core content that is evidence-based		Deep conceptual and procedural knowledge	x	Reflective teaching in a collaborative setting
	x	Teaching for understanding	x	Rigorous curriculum		Development of global perspective
	x	Teaching through problem solving and inquiry	x	Relevant curriculum		Incorporation of current technology
		Student-centered classrooms	x	Assessment for learning		
Cognitive Domain		Remembering	x	Applying		Evaluating
	x	Understanding	x	Analyzing		Creating
Connections to Students' Lives		Isolated within discipline	x	Connected to other disciplines		Connected to student lives
	x	Connected within discipline		Has value beyond school purposes		
Support for Literacy		Using the literacy process for inquiry	x	Increasing reading volume		
		Increasing access to print	x	Engaging students with texts		
		Involving students in discussion		Reading aloud in content areas		
	x	Increasing reading fluency		Explicitly instructing in vocabulary and comprehension		
		Writing to learn across content areas				
Class Time	1 class period(s) 50 total minutes					

Materials & Set-up:
Statistical software such as Fathom, student page, access to a newspaper database

Teaching Tips:

After the introductory problems, the teacher will give a mini-lesson on association and causation. The teacher may want to have unusual data sets that contain odd correlation of interest to his/her students. Emphasis should be placed on the fact that a strong association does not imply causation. The remainder of the lesson is student-driven as they will search for articles. The teacher's role is to help the students

© 2008 Iowa Department of Education

stay focused on finding good examples of correlation and reading these with a critical eye. You may want to encourage partnerships to share articles and analysis before this is done as a whole class. The teacher should also come prepared with some ideas of where to find data sets. Some good places are www.fedstats.gov , www.census.gov, and the Center for Disease Control. There are many other good places to look, but these might be good starting points. For the example of causation and correlation another suggestion is from *Statistics for the Utterly Confused* p. 88. It describes an increase in the number of the skiing accidents when the number of hot chocolate sales increases. Of course, one cannot "cause" the other, but the association indicates the clear additional factor: more skiing occurs in cold weather and cold weather produces increased hot chocolate sales.

Teacher Tasks	Student Tasks
• Facilitate class discussion about readings. • Facilitate and provide mini-lecture on association and causation. • Monitor students' progress toward finding good example articles in newspaper. • Provide resources for statistical (reliable) databases on line. • Facilitate sharing session to bring the most math out possible that people are exposed to everyday. • Emphasize reading with a critical eye. • Promote participation by all students.	• Complete the reading tasks thoughtfully. • Share ideas about association and causation. • Find article including correlation and analyze it. • Learn how to use statistical data bases to find more information. • Share article, analysis and statistical information.

Differentiation Tasks

Less-Than-Proficient

Prioritized vocabulary, skills, concepts: association, causation,

Suggestions for supplemental instruction

- Partner with a strong reader
- Have articles scanned on Kurzweil
- Allow weaker or slower readers have the article in advance to pre-read during their individual skills remediation time.

Highly Proficient

Suggestions for supplemental instruction

- Students can alternatively look in a scholarly journal for an article
- Possibly investigate the data for shorter time internal such as monthly. Statistics may be available from the Des Moines Police office. Perhaps a stronger case for or against casualty vs. association might be made.

© 2008 Iowa Department of Education

Assessment:

Formative Assessments:

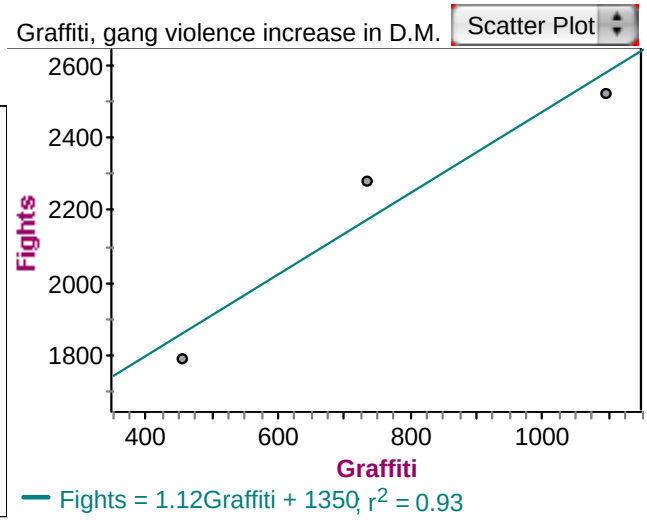
- Facilitate class discussion about readings
- Facilitate and provide mini-lecture on association and causation
- Monitor students progress toward finding good example articles in newspaper
- Solutions to the activity questions are provided.
- Facilitate sharing session to bring the most math out as possible that people are exposed to everyday.
- Promote participation by all students

1. It implies that there is a strong association between graffiti occurrences and gang violence. There is not much data included, so it would be helpful to find more data to really find the strength of the association.

2. It would be better to find some other data as well (more years).

Graffiti, gang violence increase in D.M.

	Graffiti	Fights	<new>
1	453	1793	
2	733	2280	
3	1093	2521	



Graffiti, gang violence increasÉ

	Graffiti
Fights	0.96516649

S1 = correlation ( )

The slope is 1.12. This means for each year that one more occurrence of graffiti is recorded approximately 1.12 more fights occur. The y-intercept is 1350. This means that if there were no graffiti, there would be 1350 fights.

f) No, there is an association but not causation.

© 2008 Iowa Department of Education

g) Contact the newspaper, use crime statistics found online.

h) No, it does not mean causation.

i) Students will search for articles and share with the class

Summative Assessments :

- Provide specific suggestions for evaluating students once they have completed the activity.

Reflect:

I will need to think carefully about the sharing of articles. It served my students to share in a small group and then choose one article from this group to discuss as an entire class. Since many of the students were using the same search device, some students had the same article. This was not helpful, and I will be careful to not have the same article for most students, although it may be helpful to purposely have the same article for some students (either enrichment or support).

Attachments:

© 2008 Iowa Department of Education

Activity 5
Student Page/Handout

What's in the news?

Exploration

A quick search the keyword "correlation" of a newspaper database can produce multiple articles. Reproduced on the next page is one such article printed on May 31, 2007. Read the article and answer the following questions.

1. Does the article use a correct mathematical interpretation of correlation? Why or why not?

2. Do you believe there is enough mathematical evidence to make the conclusions in the article? Why or why not?

3. Graph the data and estimate a line of best fit. Calculate the equation (without technology you may want to refer to Activity 3 #6 for help) for your line of best fit and compare it with the least squares regression equation found with technology for the data given. Interpret the slope and the y-intercept using the context of the article.

4. Do changes in the explanatory variable cause changes in the response variable? Why or why not?

5. How would you go about finding further information and data on this topic?

6. Does a strong association mean causation? Explain your answer to this question after reading the following portion of an online article and think about the relationships discussed so far, does

© 2008 Iowa Department of Education

Prediction or Coincidence?

The recent U.S. national election brought up an interesting statistical coincidence that had held true for 64 years. From 1936 through 2000, the success or failure of the NFL Redskins football team on the Sunday before Tuesday's presidential election seemed to predict the next president of the United States. If the Redskins won, the incumbent's party retained the White House. If the Redskins lost, the incumbent's party lost. Every time. Although there had been a perfect correlation for the last 17 presidential contests, the 2004 election proved this "predictor" to be flawed.

With the Redskins' loss to the Green Bay Packers on Sunday and corroborating exit polls on Tuesday, many an American voter began to wonder if the predictor would, indeed, extend its streak to 70 years. Cut to November 3rd: President Bush was settling in for another term in Washington, not packing for Crawford, Texas. For the first time in the history of the Washington Redskins, they failed to predict the winner. Astonishing!

http://www.csidata.com/techjournal/csinews/200412/page01.html#2

7. Complete a search of a newspaper database for an article using correlation or regression. Answer questions #1-4 for your article (if possible). Be ready to share your findings with the class.

© 2008 Iowa Department of Education

Paper: Des Moines Register, The (IA)
Title: Graffiti, gang violence increase in D.M.
Date: May 31, 2007

By TOM ALEX
REGISTER STAFF WRITER
Gunshot reports, fights, shootings and gang-related graffiti are on the rise in Des Moines neighborhoods. Police say they are related to each other and are an indication of the growth of youth gangs in the city. "You can show a correlation between gang graffiti and acts of violence," said Police Detective Michael Stueckrath. The reason, he said, is that graffiti often amounts to a challenge to fight. As the city goes into the warm-weather crime season, police are working on a plan to slow gang-related violence.
A police report - one of many like it in the past few days - illustrates the relationship. On May 15, a 15-year-old Hiatt Middle School student was attacked by rival gang members in a park near North High School, according to the report. The boy told police he'd been a member of the gang Asian Pride Family for about one week. Police said the boy reluctantly admitted that his friends were crossing out rival gang graffiti just before the assault. The following afternoon a 17-year-old high school student was assaulted in a southwest Des Moines neighborhood. A gunman jumped out of a car and yelled, "Don't be acting like a gangster. I'm the gangster here." Police said the gunman put the firearm to the teenager's head and ordered him to get on the ground. When the teen refused, the assailant pushed him down, kicked him in the ribs and punched him in the face. Officers said the two had clashed before, but not over gang graffiti. In this case they were fighting over arm gestures, which according to police, indicates a desire to fight. "No place is immune from it," said Police Intelligence Sgt. John Meeker. "We get reports of shots fired, gang graffiti and fights all over the city." However, the fights and gunshots seem to follow the graffiti. When he makes presentations to civic and neighborhood groups, Meeker brings maps with him to make a point. "We'll take a map of all the incidents of graffiti, a map of shots fired and a map of fights, and we overlay them," he said. "We can show that they occur in basically the same hot spots." Reports of graffiti increased 141 percent from 453 incidents reported in 2004 to 1,093 reports in 2006. During the same period, reports of fights escalated 41 percent and reports of shots fired went up 25 percent. If crime statistics continue as they have been in Des Moines, callers will report gunshots more than 600 times this year. More than 30 people will be shot. "We are well aware of the connection between gang graffiti and violence," said Marylyn Bruce, president of the Capitol Park Neighborhood Association. "That's why we try to get rid of it as quickly as possible." The city budgeted $47,000 this year to erase or paint over graffiti because, left alone, it signals a challenge to other gangs or signals disrespect.
Officers plan to beef up enforcement in downtown Des Moines in June, July and August by using school resource officers, detectives and narcotics officers all working in the same unit. It's not a gang unit, but it will work on gang crime as needed, said Police Maj. Richard Singleton. Des Moines has resisted reorganizing a gang unit since disbanding its former gang squad at the end of the 1990s. But it may have to rethink that position as other cities, like Topeka, Kan., form new gang

© 2008 Iowa Department of Education

units, police said. "If other cities start gang units again, some of their gang members may be coming here," said Singleton. Topeka Police Sgt. Ron Gish said the Topeka Police Department could no longer maintain a gang unit after Sept. 11, 2001. The department lost officers to overseas military service at a time when homeland security became an issue locally. But by 2005 Topeka police decided gang-related crime was occurring at a rate that required the attention of a special unit. Gish said Topeka police have identified 10 major gangs and some 1,000 gang members operating in the city of some 130,000. Des Moines police estimate about 30 gangs are operating here. Officers will not guess at how many gang members there are in those groups. Minneapolis and St. Paul police have a gang unit called the Metro Gang Strike Task Force. And now Hennepin County has formed the Hennepin County Violent Offenders Task Force. It is expected to put even more pressure on gang members in the Twin Cities. "That may force some of them here the way it did back in the 1990s," said Des Moines' Singleton.

"We have considerable gang activity going on in Des Moines," said Meeker. "They are not showing as much violence as they were 15 years ago but it's a tenuous situation. That can change in a moment. The level of violence can spark up."

Reporter Tom Alex can be reached at (515) 284-8088 or talex@dmreg.com

Graffiti and crime reports increase

Des Moines police say that reports of gang-related graffiti and fights and gun violence have gone up over the past three years. The two types of crimes are related, which is why the city will spend $47,000 to erase graffiti this year.

GUNSHOTS REPORTED:
482in 2004
582 in 2005
605 in 2006

FIGHTS:
1,793in 2004
2,280in 2005
2,521 in 2006

GRAFFITI COMPLAINTS:
453in 2004
733in 2005
1,093 in 2006

Graphic Map and Statistic Source: Des Moines Police Department/THE REGISTER - Graffiti and crime increase: Des Moines police say that reports of gang-related graffiti, fights and gun violence have gone up over the past three years. Graffiti and violence are related, which is why the city will spend $47,000 to erase graffiti this year.

Photo By: Andrea Melendez/The Register

Gurbachan Bedi, owner of Bedi Mart, 2917 Payne Road, has alerted police to new graffiti outside his store. Bedi says one of his neighbors has put up video cameras to catch the people spraying graffiti. Des Moines has budgeted $47,000 this year to erase or paint over graffiti. Police say the markings signal a challenge to other gangs or signal disrespect.

© 2008 Iowa Department of Education

Copyright © The Des Moines Register. All rights reserved. Reproduced with the permission of Gannett Co., Inc. by NewsBank, inc.
Author: Alex Tom
Section: Metro Iowa
Page: 1B
Copyright © The Des Moines Register. All rights reserved. Reproduced with the permission of Gannett Co., Inc. by NewsBank, inc.

ACTIVITY 6

Activity 6 Overview: *House Price vs. Size*

Iowa Core Curriculum Essential Skills/Concepts:	o Understands and applies descriptive statistics o Organizes and consolidates his/her mathematical thinking through communication o Uses the language of mathematics to express mathematical ideas precisely o Recognizes and applies mathematics in contexts outside of mathematics o Uses representations to model and interpret physical, social, and mathematical phenomena					
Big Ideas	o Understands appropriate uses of statistics in the unit o Models data appropriately o Use statistical reasoning o Communicates mathematical understanding effectively including the use of tables and graphs					
Characteristics of Instructional Core	x	Important core content that is evidence-based	x	Deep conceptual and procedural knowledge	x	Reflective teaching in a collaborative setting
	x	Teaching for understanding		Rigorous curriculum		Development of global perspective
	x	Teaching through problem solving and inquiry	x	Relevant curriculum	x	Incorporation of current technology
	x	Student-centered classrooms	x	Assessment for learning		
Cognitive Domain		Remembering	x	Applying		Evaluating
		Understanding	x	Analyzing	x	Creating
Connections to Students' Lives		Isolated within discipline	x	Connected to other disciplines	x	Connected to student lives
		Connected within discipline		Has value beyond school purposes		
Support for Literacy		Using the literacy process for inquiry		Increasing reading volume		
		Increasing access to print		Engaging students with texts		
		Involving students in discussion		Reading aloud in content areas		
		Increasing reading fluency		Explicitly instructing in vocabulary and comprehension		
		Writing to learn across content areas				
Class Time	2 class period(s) 100 total minutes					

Materials & Set-up:
Statistical software such as Fathom, access to the Internet, computers, student page

Teaching Tips:
It is helpful to show students how to access the information from different real estate or assessor's websites. This project is a one where students are typically able to work independently. Teachers may want to give students choice to investigate the relationship between any two variables or several variables

© 2008 Iowa Department of Education

as a replacement to the unit. If students opt for this choice, all the mathematical elements of the project should be maintained. If time allows, sharing the findings can be interesting. This may be done as a gallery walk. Of course, this project should be adapted for the community where the students live.

Teacher Tasks	Student Tasks
• Launch the lesson by inquiring about size of homes, apartments, etc. • Is this true anywhere? Why or why not? • Demonstrate how to access the data needed for the project • Allow students time to get started on project. Answer clarifying questions. • Assign the project for students to complete outside of class. • Have a day to ask clarifying questions • Encourage sharing of the information	• Participate in launch • Choose the location • Begin gathering data • Analyze data and used statistical reasoning • Create written summary

Differentiation Tasks

Less-Than-Proficient

Suggestions for supplemental instruction

- Do preliminary search and bookmark some location with rich data possibilities to allow more focus to be on analysis (support)
- If co-teaching, this allows more support to track progress for the project. (support)

Highly Proficient

Suggestions for supplemental instruction

- Allow students to choose multiple variables to research instead of the home assignment.
- Since location of the home will also influence the price, a house of fixed size can vary dramatically from one location to another, even in the same county. Have students investigate this topic and consider smaller or larger regions for the data sets. For larger regions, for example, one might try to find an average set of space-price data points for the entire state of Iowa. For smaller regions students could find the most and least expensive places (for fixed house size, say) to live within a given county or city, and investigate reasons for that.

Assessment:

Formative Assessments:

- Allow students time to get started on project. Answer clarifying questions
- Have a day to ask clarifying questions
- Encourage sharing of the information

© 2008 Iowa Department of Education

See attached rubric.

Summative Assessments:

- Provide specific suggestions for evaluating students once they have completed the activity.

Reflect:

The student activity sheet may provide too much scaffolding for this to be truly a quadrant D activity. I am thinking about creating two versions of this activity. One version will have more open-ended questions and one with the current questions. Students found this activity engaging, but also became too inquisitive about neighbors and other student's house prices. In addition it is easy to get distracted by exploring lots of interesting data and not get to the analysis. I will be sure to discuss this at the beginning of the class with the hopes of curbing some of this "play" with the online resources. If students do not have access to a computer at home, they will need structured time at school to complete the project or they may need the data supplied to them. Some students may need help in prioritizing tasks and identifying project components that can be completed at home and those that can't be.

Attachments:

© 2008 Iowa Department of Education

Activity 6
Student Page/Handout

House Size vs Price?

Data Analysis Project: House Size vs. Price?

In this project you will compare the size of a home and its selling price (or tax evaluation price) in the Cedar Valley and from a different community. You should seek answers to the following overarching questions:

**Is there a correlation between size of a home and selling price in the Cedar Valley (or your neighborhood)?*
**Is there a correlation between size of a home and selling price in the place of your future home?*
**Are the models you created for the Cedar Valley and future home comparable? What do they indicate about the area?*
**What interpretations can be made using statistical tools about size and selling price of homes?*

Requirements of Project:

1. Find the size (in hundreds of sq. ft.) and selling price (in thousands of dollars) of at least 10 homes. Choose from the following options:
   a. Within 1 mile of your own home.
   b. Within 1 mile of NU.

   You can go to http://www.co.blackhawk.ia.us/depts/bhentry.htm to find this information.

   i) Create a mathematical model relating size and selling price.
   ii) Explain the meaning of your model completely for your situation. Using statistical tools, evaluate whether size is a useful predictor of price in your area.

2. Choose a location you hope to live in the future. (If you want to remain in the Cedar Valley choose a different community that you enjoyed visiting.) Find the size and selling price of at least 10 homes in a specific neighborhood. You can go to www.realtor.com to help you find selling prices. Use the map feature of this website to be sure the homes are close to each other.

   i) Create a mathematical model relating size and selling price.
   ii) Explain the meaning of your model completely for your situation. Using statistical tools, evaluate whether size is a useful predictor of price in your area.

3. How can you use your mathematical model to estimate the price of a home in the neighborhood of your choice for Part 1?

   i) Choose a home in the neighborhood of your choice for Part 1, find its size and calculate the predicted value of the home? (This can be your own if you wish.)
   ii) Choose a home from your future neighborhood. Find its size and calculate the predicted value of the home?
   iii) Calculate the residuals or errors in prediction of your home values in (i) and (ii).

© 2008 Iowa Department of Education

4. In a paragraph, compare and contrast the size vs. value of the Cedar Valley neighborhood and your future home. Use statistical reasoning to justify any statements you make.

© 2008 Iowa Department of Education

Rubric **Data Analysis Project: House size vs. Price?**

	Must resubmit	D (Should revise)	C (Should revise)	B (Satisfactory Response)	A (Expected Response)
Cedar Valley Model (x 2)	Incomplete work	Partially complete. Did not use at least 10 data points. Inaccurate information.	Mostly complete and accurate model and graph. Explanation is lacking mathematical vocabulary and/or statistical reasoning	Mostly accurate, complete and thorough model, graph and explanation using appropriate mathematical language and constructs. Uses statistical reasoning in explanations.	Accurate, complete and thorough model, graph and explanation using appropriate mathematical language and constructs. Uses statistical reasoning in explanations.
Future Home Model (x 2)	Incomplete work	Partially complete. Did not use at least 10 data points. Inaccurate information.	Mostly complete and accurate model and graph. Explanation is lacking mathematical vocabulary and/or statistical reasoning	Mostly accurate, complete and thorough model, graph and explanation using appropriate mathematical language and constructs. Uses statistical reasoning in explanations.	Accurate, complete and thorough model, graph and explanation using appropriate mathematical language and constructs. Uses statistical reasoning in explanations.
Predicted Value (x 1)	Not all calculations completed	Calculations complete but contain substantial errors.	Computes home value but response contains multiple errors. Does not accurately identify residual/error of prediction	Computes home value using equations with little error. Accurately identifies and calculates residual/error of prediction.	Computes home value using equations with no error. Accurately identifies and calculates residual/error of prediction.
Compare and Contrast (x 2)	Incomplete work	Interpretation is complete, but demonstrates limited understanding of the meaning of statistical tools. Written explanation is unclear or confusing.	Interprets data accurately. Written explanation is understandable and clear, but contains incomplete or limited justification for statements.	Most statements are justified by logical statistical reasoning. Interprets data accurately. Written explanation is understandable and clear.	Defends each statement with logical statistical reasoning. Interprets data accurately. Written explanation is understandable and clear.
Presenta-tion of paper (x 1)	Not legible or professional in appearance	Difficult to follow, not labeled. Handwritten. Uses formal and professional language.	Not typed. May be unclear. Does not use formal and professional language.	Typed and understandable. Uses formal and professional language.	Neat, typed, easy to understand. All data included. Graphs are labeled. Uses formal and professional language.

© 2008 Iowa Department of Education