SECTION 5

EXPLORING RELATIONSHIP

Investigation 5.1
Do Names and Cost Relate?

Overview

In Investigation 3.3, "How Expensive Is Your Name?" students analyzed the statistical question, "How expensive is it to monogram their first names onto a T-shirt?" Instead of using a constant cost per letter, as was done there, the cost of a letter is based on its frequency of use in the English language in this investigation. Students will investigate the concept of any positive and negative **relationship** between two **quantitative** variables by constructing **scatterplots** of length of first name and cost of first name, as well as cost of first and last names.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level B activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Graph ordered pairs on a scatterplot
- Interpret data presented in a scatterplot

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

Common Core State Standards Grade Level Content

8.SP.1 Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities.

NCTM Principles and Standards for School Mathematics

Algebra

Grades 3–5 All students should represent and analyze patterns and functions, using words, tables and graphs; model problem situations with objects and use representations such as graphs, tables, and equations to draw conclusions.

Grades 6–8 All students should relate and compare different forms of representation for a relationship.

Data Analysis and Probability

Grades 3–5 All students should propose and justify conclusions and predictions that are based on data; represent data using tables and graphs.

Grades 6–8 All students should make conjectures about possible relationships between two characteristics of a sample based on scatterplots of the data.

Materials

- Table 5.1.1 Occurrence of Letter Percentages (available on the CD)
- Table 5.1.2 Cost of Letters (available on the CD)
- Data collection sheet (available on the CD)
- Small sticky notes (two for each student)
- Enough board space for two large graphs

Estimated Time

1–2 days

Instructional Plan

✏ Formulate a Statistical Question

1. Before creating a statistical question to investigate, spend some time having your students understand how costs could be assigned to letters based on the frequency of occurrence of letters in the English language. Ask your students which letter of our alphabet they believe is used most often. Encourage them to activate their prior knowledge by asking them to think about playing or watching games such as Scrabble, Hangman, and Wheel of Fortune. You may have them count the letters in a page of their textbook

and create a frequency/percentage table. This investigation will use the letter percentages as shown in Table 5.1.1, which is from Wikipedia.

Table 5.1.1 Occurrence of Letter Percentages

Letter	a	b	c	d	e	f	g	h	i
Percentage	8.2%	1.4%	2.8%	4.2%	12.7%	2.2%	2.0%	6.1%	7.0%
Letter	j	k	l	m	n	o	p	q	r
Percentage	0.2%	0.8%	4.0%	2.4%	6.7%	7.5%	1.9%	0.1%	6.0%
Letter	s	t	u	v	w	x	y	z	
Percentage	6.3%	9.1%	2.7%	1.0%	2.4%	0.2%	2.0%	0.1%	

2. Have your students make some observations about the table. For example, what must be the sum of all of the percentages in the table? Why? Which letter occurs most often? Least often? Do vowels occur more often than consonants? Be sure your students understand that letters with high percentages occur more frequently and letters with low percentages occur less frequently.

3. Display Table 5.1.2 Cost of Letters. Have your students make some observations about this table. For example, which letters are the most expensive? Least expensive? How is Table 5.1.2 Cost of Letters related to Table 5.1.1 Occurrence of Letter Percentages? Note that your students should discover that the letters occurring more frequently were assigned higher monetary costs, and the letters occurring less frequently were assigned lower monetary costs. Ask your students if other assignments of costs to letters could have been made. Ask your students what the cost of names with many vowels in them might be as compared to those with few vowels.

Table 5.1.2 Cost of Letters (cents)

A	B	C	D	E	F	G	H	I	J	K	L	M
6¢	3¢	3¢	4¢	7¢	3¢	3¢	5¢	5¢	1¢	2¢	4¢	3¢
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
5¢	5¢	3¢	1¢	5¢	5¢	6¢	3¢	2¢	3¢	1¢	3¢	1¢

4. Ask your students if they think there is a relationship between the length of their first name and the cost of their first name. More specifically, have them discuss which names would be the most expensive—names with more vowels or names with more overall letters? Do they think a person who has a long first name will have an expensive first name? In addition, have them consider the comparative cost of their first and last names. Do they think a person who has an expensive first name will have an expensive last name, or perhaps an inexpensive last name? Explain that they will be looking for these kinds of relationships. Lead your students to formulate the statistical questions, "Is there a relationship between the length of first names and the cost of first names?" and "Is there a relationship between the cost of first and last names?"

Collect Appropriate Data

1. Before collecting data, for consistency purposes, decide whether all students will use their given name or a nickname (e.g., Jennifer or Jenny; Katherine or Katie, Joseph or JP).

2. Using Table 5.1.2, ask each of your students to find the lengths and costs of their own first and last names. For example, John Smith would have a length of 4 for his first name, a cost of 16¢ for his first name, a length of 5 for his last name, and a cost of 24¢ for his last name.

3. Collect the class data on the data collection sheet as in Table 5.1.3.

Table 5.1.3 Data Collection Sheet

Student	Length of First Name	Cost of First Name (¢)	Length of Last Name	Cost of Last Name (¢)
1				
2				
3				
4				
5				
…				

4. For this investigation, the data in Table 5.1.4 that were collected from an 8th-grade class of 18 students will be used as an example.

Table 5.1.4 Sample Data of Costs of First and Last Names

Student	Length of First Name	Cost of First Name (¢)	Length of Last Name	Cost of Last Name (¢)
1	5	26	11	43
2	4	17	7	38
3	5	20	8	42
4	8	39	8	42
5	6	25	6	35
6	8	41	7	38
7	4	20	7	33
8	7	34	6	30
9	4	17	6	29
10	5	27	9	45
11	5	26	10	46
12	9	48	5	20
13	7	32	5	21
14	5	26	8	40
15	7	36	5	24
16	5	28	6	24
17	7	39	7	31
18	7	32	8	29

Analyze the Data

1. Ask your students to look at the columns Length of First Name and Cost of First Name. Tell them that they are to determine if there is a relationship between the length of the first name and its cost to monogram. Explain that a positive relationship means points on the graph go from the lower left part to the upper right part, so shorter names would be less expensive and longer names would be more expensive.

2. Explain to your students that constructing a scatterplot may help them discover any relationship between length of first name and its cost. Discuss with them how to construct a scatterplot. On the board, draw a horizontal axis labeled Length of First Name and a vertical axis labeled Cost of First Name. Discuss with your students the scales needed on both axes.

3. Give each student a small sticky note. On the sticky note, your students should write length and cost of their first name as an ordered pair. Figure 5.1.1 shows what John Smith would write on his sticky note.

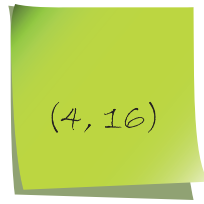


Figure 5.1.1 Sticky note with length of first name and cost of first name written as an ordered pair

4. Have your students place their sticky notes at the appropriate coordinate location on the scatterplot. Figure 5.1.2 is a scatterplot based on the 8th-grade class data. **Note:** This sample graph is drawn with points, while your class graph will have sticky notes.

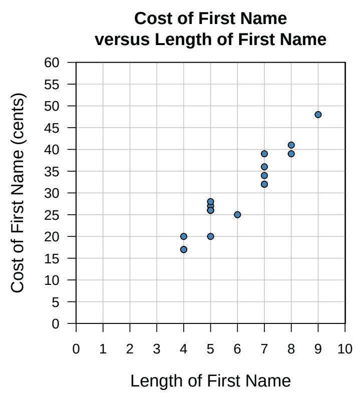


Figure 5.1.2 Scatterplot of cost (¢) of first name versus length of first name. **Note:** There are two data points at coordinate (4,17), three at (5,26), and two at (7,32).

5. Point to one of the student's sticky notes and ask your students what the coordinates of that point represent.

6. Ask your students whether they observe a pattern or relationship between the length of first name and the cost of first name. Explain that the scatterplot shows a positive trend or relationship between the length of first name and its cost. The points are generally going from the lower left part of the graph to the upper right part of the graph. Explain that this means a person who has a long first name would tend to have a more costly first name, and a person who has a short first name would tend to have a less costly first name.

7. Ask your students why they think this graph shows a positive relationship.

8. Refer to the data collection sheet (Table 5.1.4). Ask your students to look at the columns Cost of First Name and Cost of Last Name. Ask them if they see any relationship between these two variables.

9. Give students another sticky note and ask them to record the cost of their first name and the cost of their last name as an ordered pair. Figure 5.1.3 shows what John Smith would write on his sticky note.

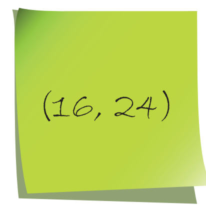


Figure 5.1.3 Sticky note of cost of first and cost of last name for John Smith written as an ordered pair

10. On the board, draw a horizontal axis labeled Cost of First Name and a vertical axis labeled Cost of Last Name. Discuss with your students the scales needed on both axes.

11. Have your students place their sticky notes at the appropriate coordinate locations on the scatterplot. Figure 5.1.4 is a scatterplot based on the 8th-grade class data. Note: This sample graph is drawn with points, while your class graph will have sticky notes.

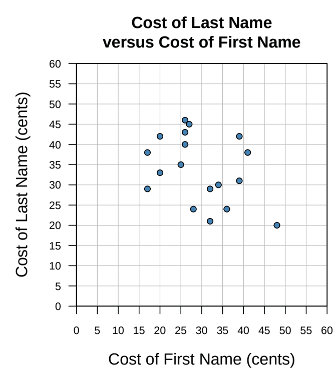


Figure 5.1.4 Scatterplot of cost (¢) of last name versus cost (¢) of first name

12. Ask your students if they observe any pattern or relationship between the cost of monogramming first names and the cost of monogramming last names. Explain to your students that the scatterplot shows a negative trend or relationship between the cost of the first and last names. The points are generally going from the upper left part of the graph to the lower right part of the graph. Explain that this means a higher cost of the first name is related with a lower cost of the last name, or a person who has an expensive first name would tend to have a less expensive last name.

13. Ask your students why they think the cost of monogramming one's first name has a negative relationship with the cost of monogramming one's last name.

✏ Interpret the Results in the Context of the Original Question

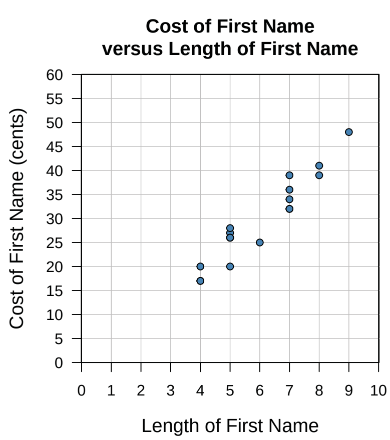
1. Remind your students of the original questions they have been investigating: "Is there a relationship between the length of first names and the cost of first names?" and "Is there a relationship between the cost of first name and cost of last name?" Have them write a summary of their investigation by answering the questions and using their graphical analysis to support their answers.

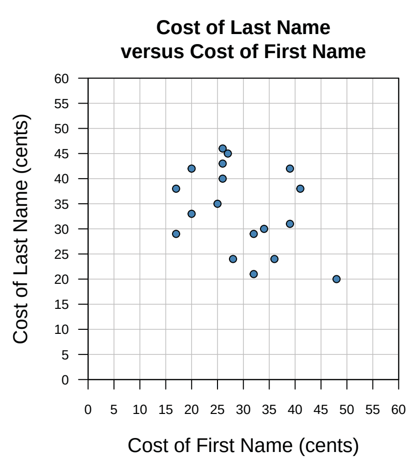
2. Ask your students if they think the cost of names using Table 5.1.2 in this investigation would be the same as for students in England? France? China? **Note:** the name John translates to Juan in Spanish and Yue Han in Chinese.

Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

In this investigation, we looked at two questions. One was on relating the length of our first names and the cost of monogramming them on T-shirts. The second question was on investigating if the cost of monogramming our first names and last names were related. The assignment of costs to letters was based on the frequency of usage of letters in English. High-frequency letters cost more and low-frequency letters cost less. In Scrabble, it's the opposite, with the letters that don't occur very often being worth more. To see how the length of our first name and its cost are related, we displayed the length and cost of our first names on a scatterplot using sticky notes. We did the same for the cost of our first and last names. Here were our graphs.





We saw that the cost of the name was higher for longer names, which meant there was a positive association between the length of our first name and its cost. We also investigated the relationship between the cost of our first and cost of our last name. We displayed a scatterplot of the cost of both names and observed that the higher costs of the first names were associated with lower costs of the last names. This meant there was a negative relationship between the cost of the first and last names.

We are going to continue this study by analyzing names in foreign countries such as China and Russia. For example, John in Chinese is Yue Han, which has a cost of 29 cents. John in English was 16 cents. Maybe Chinese names have more vowels, so they might cost more. We'll see.

Assessment with Answers

A group of 8th-grade students investigated the statistical question, "Is there a relationship between the length of their last name and the cost of their last name." Figure 5.1.5 is a scatterplot of the length of the students' last names and the cost of their last names.

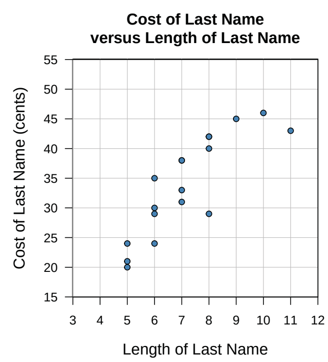


Figure 5.1.5 Scatterplot of cost (¢) of last name versus length of last name

1. Choose a point on the graph and describe what it means in the context of the variables. The point (10,46) means a last name of 10 letters would cost 46¢.

2. If a student has a long last name, does that student tend to have a more or less expensive cost for their last name? Explain your answer. A student with a long last name would tend to have a more expensive last name.

3. Overall, is there a relationship between the length of a student's last name and the cost of their last name? Use words and numbers to explain your answer. There is a positive relationship between length of last name and cost. As the length of the name increases, the cost tends to increase.

Alternative Assessment with Answers

March is National Reading Month and a teacher wanted to know if her students read more books in March than in February. Figure 5.1.6 is a scatterplot of the number of books sixth-graders each read during February and March.

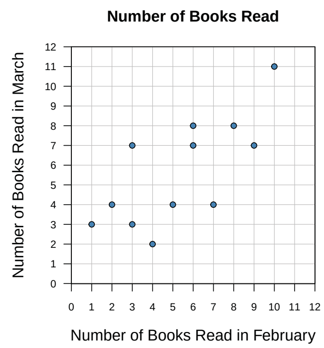


Figure 5.1.6 Scatterplot of number of books read by sixth-graders in March versus number of books read by sixth-graders in February

1. Choose a point and describe what it means in the context of the variables. Point (10,11) means this student read 10 books in February and 11 books in March.

2. If a student read many books in February, what did that student tend to do in March? Explain your answer. The student read many books in March as well.

3. Overall, is there a relationship between the number of books sixth-graders read in February and the number of books they read in March? Use words and/or numbers to explain your answer. Yes, the trend in the graph shows that the more books read in February, the more books read in March. The graph goes from the lower left to the upper right.

Extension

1. Have your students investigate the possibility of a relationship between the total length of both names and the total cost of both names.

2. The relevance of this investigation involving the frequency of letters is in the discipline of cryptography. You may want to discuss the deciphering of secret codes with your students.

References

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics, *www.corestandards.org*.

Wikipedia, *http://en.wikipedia.org/wiki/Letter_frequency*.

Investigation 5.2
How Tall Were the Ancestors of Laetoli?

Overview

The focus of this investigation is to look for and measure the degree of any **relationship** between two **quantitative** variables, specifically height and foot length. The motivation for this study comes from a science dig. Footprints were found that were determined to be more than 3 million years old. It is of interest to predict how tall those ancestors might have been based on the lengths of their footprints. Students will investigate whether there is a relationship between their own height and foot length. They will collect, organize, and analyze such data and then informally predict what the height of the ancestors might have been.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level B activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Learn to make conjectures about the relationship between two quantitative variables
- Demonstrate an ability to organize their data and display them in a scatterplot
- Learn to quantify the degree of relationship between two quantitative variables by developing the Quadrant Count Ratio (QCR)

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
6. Attend to precision.

Common Core State Standard Grade level Content

8.SP.1 Construct and interpret scatterplots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability

Grades 6-8 Students should formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population; select, create, and use appropriate graphical representations of data, including histograms, boxplots, and scatterplots.

Materials

- Class data recording sheet (available on the CD)
- Sticker dots (3/4” diameter, four colors - green, blue, yellow, and red)
- Metric sticks
- Tape
- Graph paper
- Calculators
- Laetoli background information (available on the CD)

Estimated Time

Two days

Instructional Plan

✏ Formulate a Statistical Question

1. Begin this investigation by asking your students if they think there is any relationship between the size of a person's foot and his/her height. Do people with longer feet tend to be taller? Explain to your students that scientists look for relationships like this so they can estimate the height of people who lived a long time ago. Share with your students the following background information from Wikipedia about Laetoli, Tanzania (available on the CD).

   There is a place in Tanzania, Africa, known as Laetoli. It is a special place because it is where scientists believe our ancestors of long ago walked side-by-side. It is where scientists have worked to get an understanding of the past.

   In the late 1970s, two sets of footprints were discovered at Laetoli. There were 70 footprints in two side-by-side lines 30 meters long, preserved in volcanic ash. Apparently, a volcano exploded sending ash everywhere and the two individuals just happened to walk through the area, preserving their footprints. Fossil remains in the area tell scientists that the ancestors who left the footprints found at Laetoli lived about 3.5 million years ago.

   We know the size of the feet because Dr. Mary Leakey, an anthropologist, and her team made copies of the prints using plaster casts. The locations of the footprints were put on a map, so the length of stride (distance between footprints) also can be determined. Based on these observations, foot dimensions and stride length for the two ancestors are given in Table 5.2.1. These are averages based on the 70 observed footprints.

Table 5.2.1 Footprint Data Collected by Dr. Leakey at Laetoli

	Ancestor 1	Ancestor 2
Length of Footprint	21.5 cm	18.5 cm
Width of Footprint	10 cm	8.8 cm
Length of Stride	47.2 cm	28.7

   Much has been learned from these footprints. They share many characteristics with the prints made by modern human feet.

   A research question of interest to the scientists was "How tall were these ancestors at Laetoli?" The foot length, foot width, and length of stride can be used to produce estimates of the heights of these ancestors.

2. After explaining the background information, discuss with your students how they can help the scientists answer the question, "How tall were these ancestors at Laetoli?" Tell your students they are going to focus on whether foot length and height are related. **Note:** Depending on the sensitivity of measuring feet, you may have your students measure shoe length instead. (However, to be realistic, the ancestors did not wear anything on their feet.) Although collecting real data is desirable, there is a sample set of class data given in Table 5.2.2.

3. Lead your students to formulate the statistical question, "What, if any, is the relationship between height and foot size of humans?"

✏ Collect Appropriate Data

1. Discuss with your students how they are to measure the length of their right foot. The measurements should be made in centimeters and without shoes from the back of the foot to the longest forward point of their toes. **Note:** You may want to have your students press the back of their right foot against a wall to increase the accuracy of the measurement. When measuring the height of the person, remind students they should stand straight with their back against a wall.

2. After students have measured the length of both their right foot and their height, collect the class data on the recording form (available on the CD). Table 5.2.2 is a sample set of data collected from a class of 8th-graders.

Table 5.2.2 Sample Set of 8th-Grade Class Data

Student Number	Foot Length cm	Height cm	Student Number	Foot Length cm	Height cm
1	28	175	14	24	168
2	26	181	15	23	168
3	24	168	16	23	176
4	26	168	17	27	177
5	27	178	18	25	171
6	24	174	19	22	160
7	28	179	20	27	187
8	23	157	21	28	167
9	29	190	22	27	184
10	26	170	23	29	181
11	23	169	24	27	174
12	23	166	25	22	155
13	26	174	26	24	170

✏ Analyze the Data

1. Ask your students to find the mean length of the right foot data and the mean of the height data.

2. Draw a large scatterplot on the board and plot the ordered pair (mean foot length, mean height) with a black dot. For the sample class data, the point is (25.4, 172.6).

3. Ask all students with above-average right foot length and above-average height to stand. Give them a green "sticker dot" and have them place their stickers on the graph at the appropriate coordinates.

4. Ask all students with below-average right foot length and above-average height to stand. Give them a blue "sticker dot" and have them place their stickers on the graph at the appropriate coordinates.

5. Ask all students with below-average right foot length and below-average height to stand. Give them an orange "sticker dot" and have them place their stickers on the graph at the appropriate coordinates.

6. Ask all students with above-average right foot length and below-average height to stand. Give them a red "sticker dot" and have them place their stickers on the graph at the appropriate coordinates. Figure 5.2.1 is a scatterplot of the sample set of data collected from an 8th-grade class. Note the ordered pair of means is the black dot in the scatterplot.

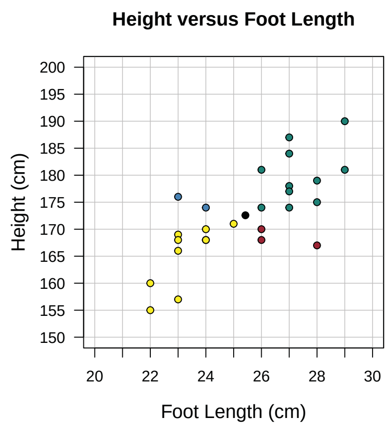


Figure 5.2.1 Class scatterplot of height versus foot length. **Note:** There is a duplicate data point at (24, 168).

7. Ask your students what trends they observe in the graph. Note that they should say there is a positive trend with longer foot lengths related to taller heights and shorter foot lengths related to shorter heights. Discuss

with your students that, in statistics, single summary numbers are calculated for a data set that tell us something about the data set. For example, when asked to characterize a central tendency for a data set, three summary statistics have been developed: the mode (most often), median (middle of the ordered data), and mean (a fair share value or balance point). Each is a single number. Similarly, when characterizing the spread of a data set, three summary statistics have been developed: the range (overall span of the data), interquartile range (span of the middle 50% of the data), and mean absolute deviation (MAD, a fair share value for how far the data are in terms of absolute distance from their mean). In this investigation of two variables, we want to develop a summary statistic (single number) that measures how related two quantitative variables are to each other. The following steps help your students develop such a summary statistic.

8. Draw a vertical line through the center point (black dot) extended to the x-axis; indicate the mean of X, 25.4 cm, on the x-axis. Similarly, draw a horizontal line through the center point (black dot) extended to the y-axis; indicate the mean of Y, 172.6 cm, on the y-axis.

9. Point to different colored dots and ask your students to explain what the dots represent in relation to the center point, given by the coordinate pair (the mean foot length, the mean height).

10. Number the four quadrants as shown in Figure 5.2.2.

    Quadrant I: Green dots

    Quadrant II: Blue dots

    Quadrant III: Orange dots

    Quadrant IV: Red dots

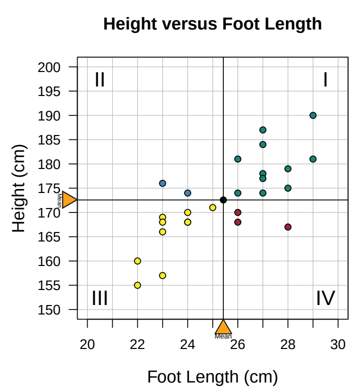


Figure 5.2.2 Scatterplot of height versus foot length showing the quadrants

The QCR

The beauty of the QCR is its simplicity and ease of calculation while providing a conceptual measure of relationship. The QCR clearly has some shortcomings. For example, the same number of points within a quadrant could occur in a very different orientation so that the data do not exhibit linearity at all and yet the QCR value would remain the same. The difficulty, of course, is that the QCR measure is based on only counts within a quadrant and does not consider, for example, how far a data point is from the horizontal and vertical axes. In high school, this topic is revisited and a more sophisticated measure (the Pearson correlation) is developed that corrects for QCR's limitations.

11. Ask your students where most of the stickers are. Determine the number of dots in each quadrant and put the number on the graph in the respective quadrants. Figure 5.2.3 shows the sample class data with the number of ordered pairs written in each quadrant. Note that there are two data points at (24,168) so there are 10 data points in quadrant III.

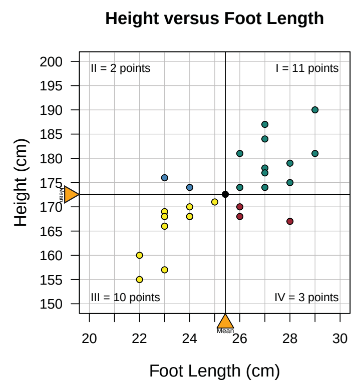


Figure 5.2.3 Scatterplot showing number of ordered pairs in each quadrant

12. Ask your students what a dot in Quadrant I represents. (People with above average foot length and above-average height). Ask what a dot in Quadrant III represents. (People with below-average foot length and below-average height)

13. Explain to your students that this graph indicates a **positive relationship** between the variables foot length and height. Generally, two numeric variables are **positively related** when above-average values of one variable tend to occur with above-average values of the other and when below-average values of one variable tend to occur with below-average values of the other. **Negative relationship** between two variables occurs when below-average values of one variable tend to occur with above-average values of the other and when above-average values of one variable tend to occur with below-average values of the other.

14. Explain to your students that we would like to have a single number that helps describe the degree of relationship seen in the graph. A **correlation coefficient** is a number that measures the direction and strength of a relationship between two variables. One such correlation coefficient is called the **Quadrant Count Ratio** (QCR). The QCR is defined as:

$$\text{QCR} = \frac{(\text{Number of Data Points in Quadrants I and III}) - (\text{Number of Data Points in Quadrants II and IV})}{\text{Total Number of Points}}$$

15. Have your students find the QCR for the class data. For the example:

$$QCR = \frac{((11 + 10) - (2 + 3))}{26} = \frac{(21 - 5)}{26} = \frac{16}{26} = 0.62$$

16. Ask your students to find the value of the QCR if all the ordered pairs are located in quadrants I and III. Ask your students if it is possible to get a QCR greater than 1 (such as 1.5). Ask your students to find the value of the QCR if all the ordered pairs are located in quadrants II and IV. Ask your students if it is possible to get a QCR less than -1 (such as -1.9).

17. Explain to your students that the closer the value is to +1, the stronger the positive relationship is. The closer to -1 suggests a stronger negative relationship. Close to 0 would indicate no relationship. Have them look at the scatterplot and explain why that should be.

18. Indicate on a number line where the class value of QCR is and ask students what the value indicates about the strength of the relationship between height and foot length. Figure 5.2.4 shows the location of the QCR for the sample set of data.

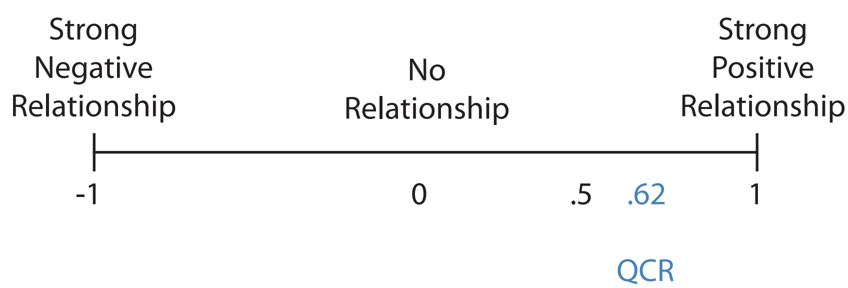


Figure 5.2.4 Strength of relationship on a number line

19. Ask your students to look at Figure 5.2.3 and, assuming the relationship of height to foot length is the same for the ancestors of Laetoli as it is in this data set, what color of sticker would the ancestors have based on a mean footprint of 21.5 cm? Note that they should say orange and that the ancestors had shorter feet than their own and were shorter in height than their own average height.

Interpret the Results in the Context of the Original Question

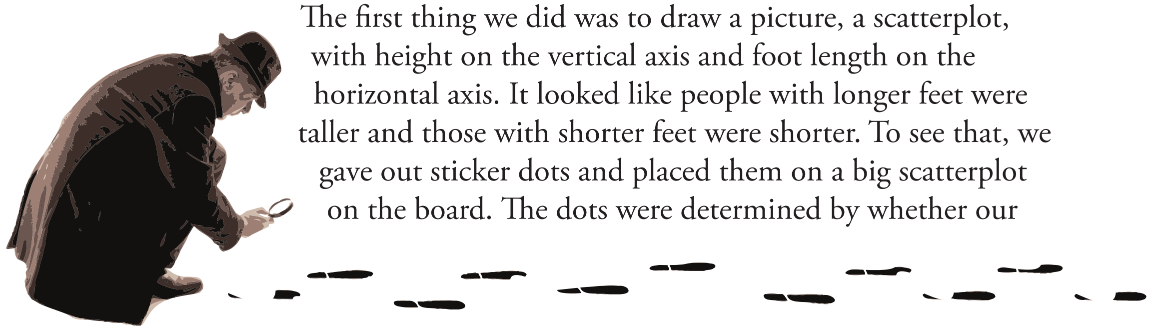
1. Have your students recall the original statistical question, "What, if any, is the relationship between height and foot size of humans?" Have your students write a brief report that answers the question and justifies their answer by using the analysis they did in class. In addition, remind your students that what prompted the statistical question involving height and foot length was an interest in trying to estimate the heights of the two ancestors from Laetoli, as we know only the length of their footprints. Have your students include in their report how they might go about coming up with an estimate of the height of the Laetoli ancestors. Indicate that you are not as interested in their actual estimate as you are in the process they are suggesting for determining the estimate. (See the extension for further development.)

2. Ask your students how they think the relationship between height and foot length would change from what they found using their class data if they collected data on height and foot length from all the teachers in the school.

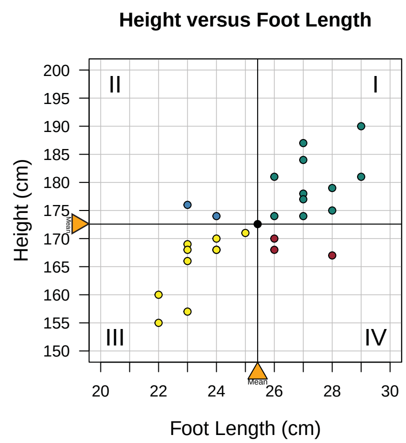
Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

For a statistics project, we got an idea from an anthropological study by Dr. Leakey, who found footprints of 3.6 million–old ancestors in Laetoli, Tanzania. The study had the ancestors footprint lengths, and we were wondering how tall they might have been. One of the set of footprints had a mean footprint of 21.5 cm. Our statistical question was, "Is there a relationship between human height and foot length?" Our data were the lengths of our right foot and our height. There were 26 paired data points in our class.



The first thing we did was to draw a picture, a scatterplot, with height on the vertical axis and foot length on the horizontal axis. It looked like people with longer feet were taller and those with shorter feet were shorter. To see that, we gave out sticker dots and placed them on a big scatterplot on the board. The dots were determined by whether our height was above or below the mean height of 172.6 cm and how our foot length compared to the mean 25.4 cm. Green dots were for (above 25.4 foot length, above 172.6 height); blue for (below, above); orange for (below, below); and red for (above, below). We added vertical and horizontal lines through the paired mean point. The scatterplot looked like this:



We could see a definite trend from the lower left to the upper right. In statistics, single numbers called summary statistics are often calculated to indicate the degree of some characteristic. So, our teacher suggested we count the number of points in the first and third quadrants and subtract the numbers in quadrants two and four, and then take the mean and call the result the Quadrant Count Ratio (QCR). For our data, QCR = ((11+10) – (2+3))/26 = 0.62. If all the data had been in quadrants one and three, the QCR would have been 1. So, we decided that .62 was pretty good and that it reflected a positive relationship. We then decided that our Laetoli ancestors would have had orange stickers, since the mean footprint we had for them was 21.5 and, from our scatterplot, there was no way the sticker could be blue. We were thinking about doing this study on all our teachers to get a new data set and see if it differs from ours. There's a difference of opinion. Some of us think it would have more variation because the ages of the teachers are more spread out than our ages.

Assessment with Answers

A group of students measured their height and arm span in centimeters. Table 5.2.3 shows the data they collected, and the scatterplot of the data is shown in Figure 5.2.5.

Table 5.2.3 Height and Arm Span (cm)

Height	Arm Span	Height	Arm Span
155	151	173	170
162	162	175	166
162	161	176	171
163	172	176	173
164	167	178	173
164	155	178	166
165	163	181	183
165	165	183	181
166	167	183	178
166	164	183	174
168	165	183	180
171	164	185	177
171	168	188	185

Figure 5.2.5 Scatterplot of arm span versus height

1. Describe the relationship between arm span and height. There is a positive relationship between arm span and height. Higher values of arm span tend to occur with higher values in height; lower values of arm span tend to occur with lower values in height.

2. Find the mean height and the mean arm span. Mean height = 172.5 cm and the mean arm span = 169.3.

3. Locate the point (mean height, mean arm span) on the graph and draw a horizontal line and a vertical line through the point.

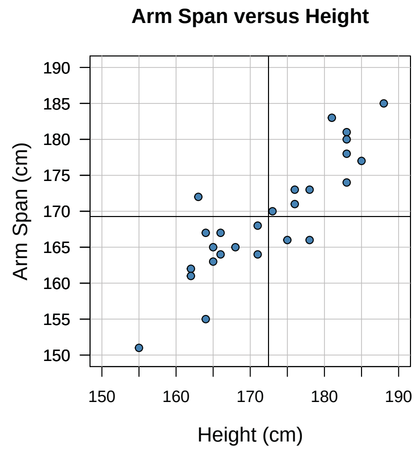


Figure 5.2.6 Scatterplot showing means and vertical and horizontal lines

4. Find the value of the QCR. QCR = ((11 + 12) - (1 + 2))/26 = (23 - 3)/26 = 20/26 = .77

5. Interpret the value of the QCR. Fairly strong positive relationship between height and arm span. This indicates that height is a pretty good predictor for arm span.

Extension

To determine an estimate for the height of the Laetoli ancestors, suggest the following:

1. Consider the scatterplot of height versus foot length as shown in Figure 5.2.7. Hand out a copy of this scatterplot to your students.

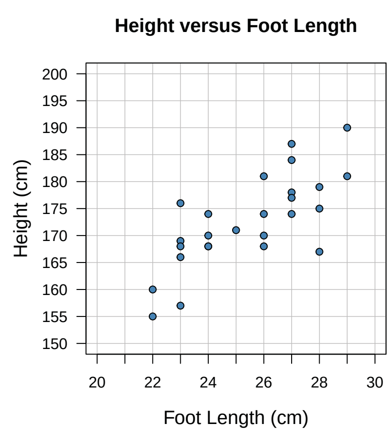


Figure 5.2.7 Scatterplot of height versus foot length. **Note:** There is a duplicate data point at (24,168).

2. Demonstrate and discuss an eyeball line on the class scatterplot using a piece of string or yarn. Ask your students what property the line should have. Lead them to suggesting that the line should "fit" the data fairly well.

3. Draw the line on the class graph and demonstrate making a prediction using the line. For a given value of X—for example, 21.5 for Laetoli ancestor 1—from 21.5 on the X axis, move vertically up to the line, then horizontally to the y-axis, as shown in Figure 5.2.8.

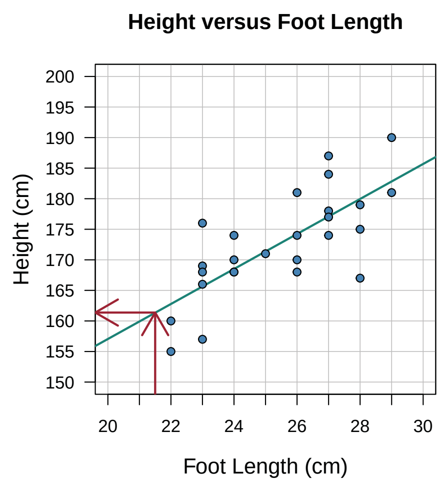


Figure 5.2.8 Scatterplot of height versus foot length with eyeball fit line. **Note:** There is a duplicate data point at (24,168).

4. Have each student draw their eyeball fit line and use the line to make a prediction for the two Laetoli ancestors.

References

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics, *www.corestandards.org*.

Laetoli, *http://en.wikipedia.org/wiki/Laetoli.*

Investigation 5.3
How Long Does It Take to Perform the Wave?

Overview

The focus of this investigation is looking for a **relationship** between two **quantitative** variables. Specifically, students will investigate whether there is a relationship between number of people and how long it takes them to perform the "wave." As part of this investigation, students will collect, organize, and analyze data by conducting an experiment to time how long it takes a varying number of students to perform the wave. Students will construct a **scatterplot**, use the plot to look for patterns in the data, and draw a line to summarize the data.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level B activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Construct a scatterplot
- Describe a relationship between two variables
- Draw a line and describe the rate of change

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

Common Core State Standards Grade Level Content

8.SP1 Construct and interpret scatterplots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

8.SP2 Know that straight lines are widely used to model relationships between two quantitative variables. For scatterplots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability

Grades 6-8 Students should formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population; select, create, and use appropriate graphical representations of data, including histograms, boxplots, and scatterplots.

Materials

- Stopwatch
- Graph paper for each student
- Ruler or straightedge for each student
- Data collection sheet (available on the CD)

Estimated Time

One day

Instructional Plan

✏ Formulate a Statistical Question

1. Begin this investigation by asking your students if they have ever been at a sporting event where the crowd performed the wave. Note that you may wish to show a YouTube video of spectators at a sporting event performing the wave. Share with your students that some people claim the wave was first performed at Fenway Park in Boston. Others claim it originated

at Pacific Lutheran University in the early 1960s. But no matter where it started, it occurs at many sporting events. Ask your students how they might predict how long it takes to perform the wave in a large football stadium. Ask your students what they would need to know to answer this question. Students usually suggest the number of people, number of sections in the stadium, and how fast the wave was performed.

2. After discussing your students' ideas, lead them to the statistical question, "Is there a relationship between number of people and length of time it takes them to perform the wave?"

Collect Appropriate Data

1. Before collecting data, discuss with your students how they are to perform the wave. Suggest they remain seated, but push their chair away from their desk. To perform the wave, they are to stand while raising their arms straight up in the air over their head, and sit back down. Have one student demonstrate the wave. Also, have your students agree on how fast they are to perform the wave. Should each student perform the wave as fast as possible or be deliberate in the motion? It is recommended to have your students be deliberate in the wave motion. It also is recommended that each student practice performing the wave to keep the procedure the same.

2. Appoint a timekeeper. The same person should do all the timing.

3. Start with three students as your first group to perform the wave. When the timekeeper says "go," the first student stands, moves his/her arms, then sits down. As soon as the first person sits down, the second student starts the wave. When the second student sits down, the third student performs the wave. Record the number of students who performed the wave and the time elapsed in Table 5.3.1.

Table 5.3.1 Data Collection Sheet

Number of Students	Time (sec) to Complete the Wave
3	
6	
9	
…	

4. Add three more students and have all six students perform the wave. Again, record the results. Continue to add three students until the entire

class has been included in performing the wave. Table 5.3.2 shows the results of 24 8th-graders having performed the wave experiment.

Table 5.3.2 Results of the Wave Experiment for a Group of 8th-Graders

Number of Students	Time (sec)
3	4
6	8
9	13
12	17
15	20
18	24
21	27
24	30

Analyze the Data

1. Ask your students if they see any patterns in the table. Students should recognize that the number of students increased by three and the time increase varied by 3, 4, or 5 seconds. To help the students focus on the change in time for the wave, add a column to the data collection sheet labeled Change in Time. Table 5.3.3 shows how the time increased as the number of students increased. Ask your students why they think the change in time varied.

Table 5.3.3 Change in Time

Number of Students	Time (sec)	Change in Time
3	4	
6	8	4
9	13	5
12	17	4
15	20	3
18	24	4
21	27	3
24	30	3

2. Ask your students to find the median number of seconds by which the change in time increased. In the example, the median is 4 seconds.

3. Ask your students about how much longer it takes to perform the wave for every additional three people. See if they realize you are asking them for a rate of change, or slope if the data turn out to be linear. For this

example, the median increase is 4 seconds for three people, or 4/3 second increase per person.

4. Ask your students how long it would take all the students in grades 6, 7, and 8 to perform the wave based on the 4/3 second per person estimate.

5. Explain to your students that a scatterplot is also useful in finding patterns or relationships between the number of students and the time to perform the wave. Have your students make a scatterplot on their graph paper. Put Number of Students on the horizontal axis (x-axis) and Length of Time on the vertical (or y) axis. Plot the ordered pairs (number of students, time). Figure 5.3.1 is a scatterplot for the example 8th-grade class data.

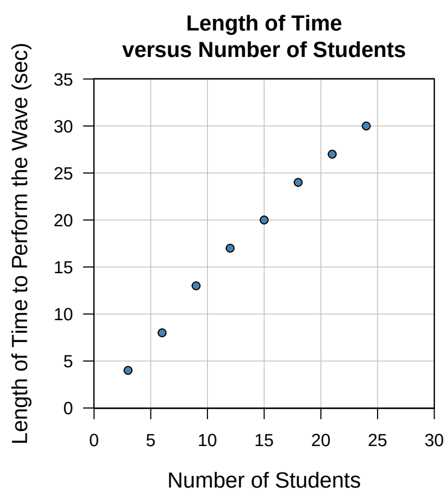


Figure 5.3.1 Scatterplot of length of time versus number of students

6. Have your students examine the scatterplot. Ask your students to describe what they observe from the scatterplot. Ask what type of relationship there is between the number of people and the length of time. And how strong is this relationship? Note that you may wish to have your students find the QCR (Quadrant Count Ratio) described in Investigation 5.2.

7. Explain to your students that you would like for them to draw a straight line through the data matching the pattern in the data as closely as they can. This line will be used to help look for patterns and make predictions about how long the wave takes. Ask them for criteria to use for determining their line. Ask them to justify why they want their line to go through or not go through the origin (0,0). Have them use a straightedge or ruler to draw the line. Figure 5.3.2 shows an example of a line drawn through (0,0) on the scatterplot from the example 8th-grade class data.

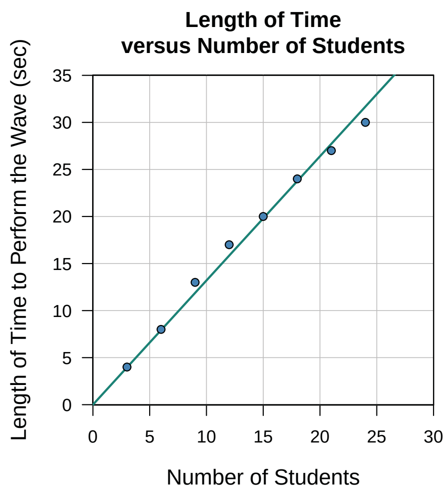


Figure 5.3.2 Scatterplot with line drawn through the (0,0)

8. Have your students locate a point on the line they drew. For this eyeball line example, a point on the line is (20,26). Ask your students what the coordinates of the ordered pair they listed represent.

9. Using the line that was drawn in Figure 5.3.2, ask your students to describe how much longer it takes to perform the wave for each additional person added. Remind them of the question you asked them in Step 3 of the Analyze the Data section. Tell your students that this value is called a **rate of change** or the **slope** of their line. For this example, the rate of change is 26/20, or about 1.3, which means that for every additional person, the time to perform the wave will increase by about 1.3 seconds.

Interpret the Results in the Context of the Original Question

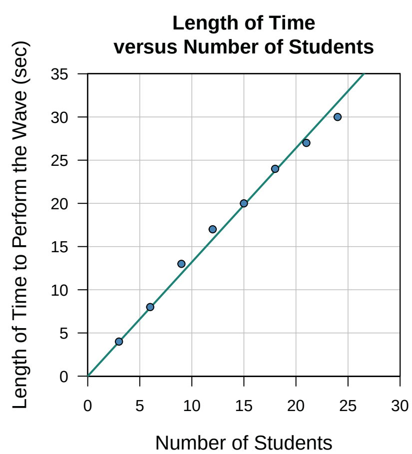
Have your students recall the original statistical question: "Is there a relationship between the number of people and the length of time to perform the wave?" Have your students answer this question in a paragraph in which they support their answers in depth using the analysis they performed. In this answer, they should refer to the relationship they observed between the number of people performing the wave and the length of time to complete the wave.

Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

This activity was really fun because we got to perform the wave in class. The statistical question we came up with was "Is there a relationship between the number of people and the length of time to perform the wave?" We actually collected data in our classroom, starting with timing how long it took three of us to perform the wave.

First, we all had to practice so we were doing the procedure the same. Otherwise, we would bias our data. We also had one timekeeper maintain all the times so no bias would enter there, either. We made a data chart by increasing the number of us performing the wave by three each time and the time it took us. We calculated that it took a median increased time of 4 seconds for every three students we added, so the rate of change is an increase of 4/3 seconds for every additional person. We also figured out that if our whole grade level of 243 students lined up to perform the wave and our rate of change was accurate, it would take 243*(4/3) = 324 seconds or about 5.4 minutes to perform the wave. Wow. We showed our data in another way by graphing the points in a scatterplot. Here it is.



We eyeballed a line through the data. We decided the line should go through the origin because it made sense that if there are no people, then the time to perform the wave is 0. We calculated a rate of change by finding a point that was on our line. The point (20,26) looked like it was on our line. So, the rate of change or slope is 26/20 = 1.3, which is about what we got before for the rate of change, 4/3. This rate means that for every additional person added, the time to perform the wave goes up about 1.3 seconds.

Assessment with Answers

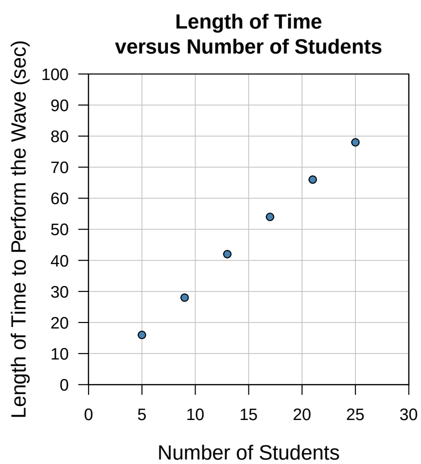
A group of 8th-grade students wanted to investigate the relationship between how long it takes to perform the wave and the number of people participating. The table below shows the results of an experiment that students conducted.

The experiment started with a group of five students. The timer said "Go" and the five students made a wave. The first student stood up, threw his/her hands in the air, turned around, and sat down. The second student did the same, and so on. The last student said "Stop" when he/she sat down. The timer recorded the elapsed time in seconds. The experiment was repeated with 9, 13, 17, 21, and 25 students.

Table 5.3.4 Number of Students and Length of Time to Perform the Wave

Number of Students	Time (sec)
5	16
9	28
13	42
17	54
21	66
25	78

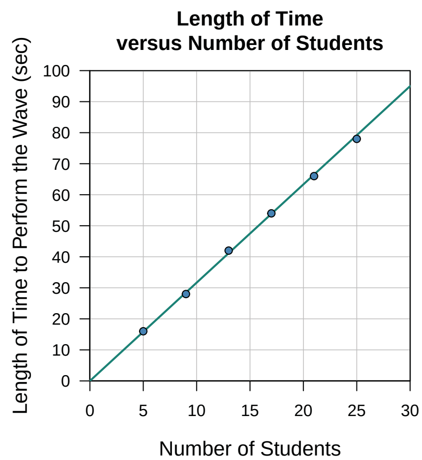
1. Draw a scatterplot of the length of time (sec) versus the number of students.



2. Is there a relationship between the number of students and the length of time to perform the wave? Describe the relationship. There is a strong positive relationship.

3. Describe any patterns you observe in the collected data for both the number of students and the length of time. As the number of students increases by four, the time to perform the wave increases by 12.4 seconds (the average of the increase changes in time).

4. Draw a line that matches the pattern in the data as closely as you can. List an ordered pair that lies on the line. Describe what the coordinates of the ordered pair represent.



The x-coordinate is the number of people and the y-coordinate is the predicted length of time to perform the wave for that number of people.

5. For each additional student added, how much longer does it take to perform the wave? Use words, numbers, and/or graphs to explain your answer. For each additional student, the wave would take a little more than 3 seconds. In the chart, the change of time for each addition of four students was about 12 seconds, which would give about 3 seconds for each additional person. The rate of change of the line on the graph is about 3.2 seconds per person.

Extension

1. Ask your students to write the equation of the line they drew through the scatterplot of Time versus Number of Students. Ask them to interpret the slope of the line in terms of the scenario.

2. Have your students predict how long it would take all the students in your school to perform the wave. Ask the principal if this could be done during an all-school program.

3. Investigate the size of a stadium near or in your community. Write down how many sections there are and how many seats are in a row. Have your students calculate how long they would predict it would take the spectators to perform the wave at the stadium.

References

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics, *www.corestandards.org*.

Wikipedia, *http://en.wikipedia.org/wiki/wave_(audience)*.

Investigation 5.4
How Do Events Change Over Time?

Overview

Analyzing trends in events that occur over time is important for people who work in government, business, and industry—and for people who study climate, the environment, or agriculture. This investigation explores ways different events change over time and develops the mathematics necessary to describe and analyze the changes. Statistics concepts include the GAISE model and graphing with a scatterplot; mathematics concepts include finding the equation of a line and interpreting its **slope** and **intercept**, as well as **rate of change.** This investigation is based on an activity in *Exploring Linear Relations*, a module in the Data-Driven Mathematics series (1998).

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level B activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Find and interpret slope as a rate of change
- Write the equation of a line from given information
- Graph a linear equation
- Make and interpret a scatterplot over time

Common Core State Standards for Mathematical Practices

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
6. Attend to precision.

Common Core State Standard Grade Level Content

8.F.3 Interpret the equation y = mx + b as defining a linear function, whose graphics form a straight line; give examples of functions that are not linear.

8.F.4 Construct a function to model a linear relationship between two quantities. Determine the rate of change and initial value of the function from a description of a relationship or from two (x,y) values, including reading these from a table or graph; interpret the rate of change and initial value of a linear function in terms of the situation it models and its graph or a table of values.

8.SP.2 Know that straight lines are widely used to model relationships between two quantitative variables. For scatterplots that suggest a linear association, informally fit a straight line and informally assess the model fit by judging the closeness of the data points to the line.

8.SP.3 Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability Standard

Grades 6-8 All students should formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population; select, create, and use appropriate graphical representations of data, including histograms, boxplots, and scatterplots; make conjectures about possible relationships between two characteristics of a sample on the basis of scatterplots of the data and approximate lines of fit.

Materials

- Table of top three money-making movies for each year from 2000–2010 for each student (available on the CD)
- One sheet of graph paper for each student

Estimated Time

Two days

Instructional Plan

✏ Formulate a Statistical Question

1. Begin the investigation by asking students how they think the cost of

popcorn at the movies has changed over time. Share with your students that a small bag of popcorn cost $0.05 in 1929 and an average of $4.75 in 2011. Much changes over time. Post the list of questions (available on CD) below and ask students to describe the changes over time.

a. How does the cost of buying groceries change over time? How about for gasoline? Cars?

b. How do middle-school students' heights change from year to year?

c. How does the number of cell phones in use change over time?

d. How does the amount of lead in the air change over time?

e. How does the number of 13- and 14-year-olds in the United States change over time?

f. How does the total revenue for fast-food restaurants change over time?

g. How does the TV rating of viewers watching *American Idol* change over time?

h. How does the temperature in your city change from January to December?

2. After discussing the questions, tell your students they will be analyzing data to help answer the statistical question, "By how much, if any, are the mean gross receipts for movies changing over time."

Collect Appropriate Data

1. Discuss with your students that there are questions that need to be considered before collecting data. For example:

   - Are the box office receipts only for the United States, or are international receipts also included?
   - Are the data to be collected for only the highest-grossing movie of the year or, say, an average of the top 10 movies per year?
   - Should the question be for all movies, regardless of category, or restricted to a specific type, such as "action?"

   This investigation is based on receipts for the top three grossing movies for each of the years from 2000–2010 and taking the mean of these three receipts per year as data.

2. Ask your students to find the three top money-making movies for each year from 2000–2010 and present the data in a table. Have them find the mean of the three movies for each year and present their results in

a table. Data may be found at *www.imdb.com/boxoffice/alltimegross* or in Table 5.4.1. Note that the gross receipts data are in millions of dollars. The Mean column of entries shown here is for the teacher. Students are to compute it. There is an activity sheet for students on the CD.

Table 5.4.1 Top Three Money-Making Movies for Years 2000–2010 in Millions of Dollars

Year	First	Second	Third	Mean
2010	415.0 Toy Story 3	334.2 Alice in Wonderland	312.1 Iron Man 2	353.8
2009	760.5 Avatar	402.1 Transformers: Revenge of the Fallen	302.0 Harry Potter and the Half-Blood Prince	488.2
2008	533.3 The Dark Knight	318.3 Iron Man	317.0 Indiana Jones and the Kingdom of the Crystal Skull	389.5
2007	336.5 Spider-Man 3	320.7 Shrek the Third	318.8 Transformers	325.3
2006	423.0 Pirates of the Caribbean: Dead Man's Chest	250.9 Night at the Museum	244.1 Cars	306.0
2005	380.3 Star Wars: Episode III – Revenge of the Sith	291.7 The Chronicles of Narnia: The Lion, The Witch, and the Wardrobe	290.0 Harry Potter and the Goblet of Fire	320.7
2004	436.5 Shrek 2	373.4 Spider-Man 2	370.3 The Passion of the Christ	393.4
2003	377.0 The Lord of the Rings: The Return of the King	339.7 Finding Nemo	305.4 Pirates of the Caribbean: The Curse of the Black Pearl	340.7
2002	403.7 Spider-Man	340.5 The Lord of the Rings: The Two Towers	310.7 Star Wars: Episode II – Attack of the Clones	351.6
2001	317.6 Harry Potter and the Sorcerer's Stone	313.8 The Lord of the Rings: The Fellowship of the Ring	267.7 Shrek	299.7
2000	260.0 How the Grinch Stole Christmas	233.6 Cast Away	215.4 Mission: Impossible II	236.3

Analyze the Data

1. Ask your students how much money Toy Story 3 grossed in 2010.

2. Ask your students which is the highest-grossing movie in all the years listed? How much money did the movie gross?

3. Ask your students what trends they observe in the data. Hopefully, they will suggest that a graph would be helpful in answering the question.

4. Have your students construct a **scatterplot**. The scatterplot can help visualize if and to what extent there is a relationship between year and mean gross receipts. Students should place the Year variable on the horizontal axis and the Mean Gross Receipts variable on the vertical axis. Note that a scatterplot with time on the horizontal is often called a **time series** plot. The plot is shown in Figure 5.4.1. Note also that a time series plot has only one data point per year, so the graph is realistically of a single numerical variable (vertical axis).

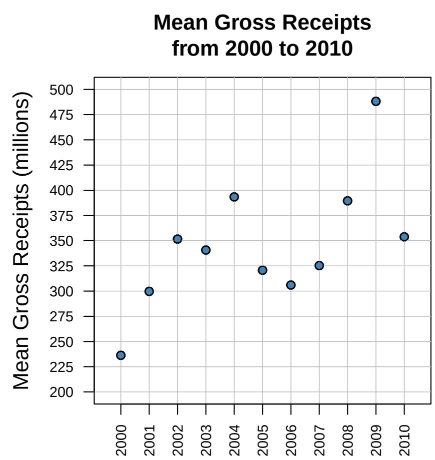


Figure 5.4.1 Time series plot of mean gross receipts (in millions) from 2000–2010

5. Ask your students if it generally appears that the mean gross receipts have increased over the years 2000–2010?

6. Ask your students if there are any years in which the mean receipts decreased from the previous year? How can these situations be identified on the graph?

7. Discuss with your students that to help visualize any increases and decreases in mean gross receipts, it would be helpful to find the change from one year to the next. Display Table 5.4.2 and explain that the change of 63.4 million dollars for the years 2000–2001 was determined by subtracting 236.3 (mean for 2000) from 299.7 (mean for 2001).

8. Ask your students how they would find the change for the years 2009–2010. Ask them why this value is a negative number.

9. Have your students complete the rest of Table 5.4.2.

Table 5.4.2 Change in Mean Gross Receipts

	2000–2001	2001–2002	2002–2003	2003–2004	2004–2005
Mean Change (millions)	+63.4				
	2005–2006	**2006–2007**	**2007–2008**	**2008–2009**	**2009–2010**
Mean Change (millions)					-134.4

The mean change for all the years is listed in Table 5.4.3.

Table 5.4.3 Change in Mean Gross Receipts (Completed Table)

	2000–2001	2001–2002	2002–2003	2003–2004	2004–2005
Mean Change (millions)	+63.4	+51.9	-10.9	+52.7	-72.7
	2005–2006	**2006–2007**	**2007–2008**	**2008–2009**	**2009–2010**
Mean Change (millions)	-14.7	+19.3	+64.2	+98.7	-134.4

10. Ask your students to find the mean of the changes in gross receipts for all years between 2000–2001 and 2009–2010. Remind students to be aware of the negative numbers when finding the sum of the changes.

11. Discuss with your students what the mean of the changes represents. **Note:** The mean represents the amount that the gross receipts would change each year if the change were a constant. In this case, the gross receipts go up an average of $11.75 million each year.

12. Show your students a different method for finding the mean of the changes by taking the overall increase in gross receipts between 2000–2010 and dividing it by 10. **Note:** (353.8 – 236.3)/10 = $11.75 million. Recall that the top three money-making movies in 2010 grossed $353.8 million and the top three money-making movies in 2000 grossed $236.3 million.

13. Discuss with your students why the two procedures result in the same value.

14. Have your students assume that mean gross receipts increased by the constant amount of $11.75 million per year. Using that assumption, ask them to estimate what the mean gross receipts in 2001 would have been. Note that 236.3 + 11.75 = 248.05 million.

15. Assuming a constant increase of 11.75 every year, complete Table 5.4.4.

Table 5.4.4 Estimated Mean Gross Receipts (Assuming a Constant Increase)

Estimated Mean Gross Receipts (Assuming a Constant Increase in Receipts from Year to Year)						
Year	2000	2001	2002	2003	2004	2005
Time	236.3					
Year	2006	2007	2008	2009	2010	
Time						

The estimated mean gross receipts assuming a constant are shown in Table 5.4.5.

Table 5.4.5 Complete Table of the Estimated Mean Gross Receipts (Assuming a Constant Increase)

Estimated Mean Gross Receipts (Assuming a Constant Increase in Receipts from Year to Year)						
Year	2000	2001	2002	2003	2004	2005
Time	236.3	248.05	259.8	271.55	283.3	295.05
Year	2006	2007	2008	2009	2010	
Time	306.8	318.55	330.3	342.05	353.8	

16. For the estimated mean gross receipts assuming a constant increase, have your students construct a scatterplot with Year on the horizontal axis and Mean Gross Receipts on the vertical axis. See Figure 5.4.2.

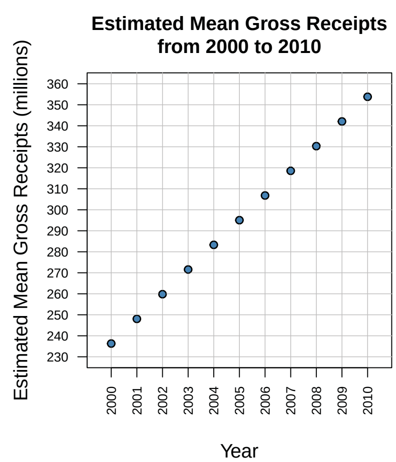


Figure 5.4.2 Time series plot of estimated mean gross receipts for 2000–2010 (assuming a constant change)

17. Ask your students to describe the pattern of the points they see in the graph. **Note:** The constant change in mean gross receipts over time is called the **rate of change.** When the rate of change is constant for equal

time intervals, the graph of the relationship is a straight line. The rate of change is also referred to as the **slope** of the line.

18. Now that your students have been reminded as to what rate of change or slope means for linear data, using the **original data** and the scatterplot of the original data, have your students choose two points so a line through them would be a fairly good prediction line in their view. For example, the line through (2001, 299.7) and (2008, 389.5) appears to fit the data well. Using those two points, discuss how to find the rate of change and how to interpret the result in the context of the problem. Then, have your students do the calculation and interpretation. **Note:** The rate of change of mean gross receipts as determined by the data from 2001–2008 is (389.5 – 299.7) / (2008 – 2001) = 89.8 / 7 = 12.83 million dollars. In words, a possible interpretation of the slope in context might be: For each one-year increase, the mean gross receipts increase by $12.83 million, on average. The time series plot with a trend line appears in Figure 5.4.3.

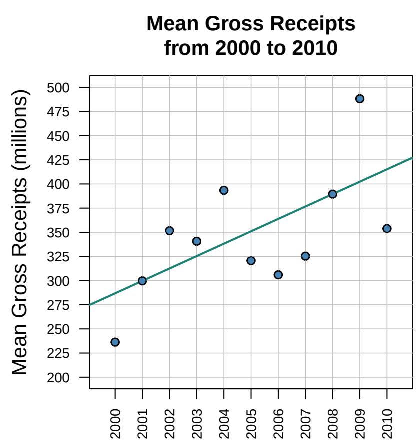


Figure 5.4.3 Time series plot of estimated mean gross receipts from 2000–2010 with trend line

19. Discuss with your students the comparison between the rate of change of 11.75 when a constant rate of change was assumed and the rate of change of 12.83 when using the original data from 2001 and 2008. This is key to this investigation. Ask your students how well they think the line drawn describes the data. Discuss what the line with slope $11.75 million is based on as compared to the line they drew with a slope of 12.83. Which is better? Note that the Common Core State Standards in statistics at the high-school level develop a more formal structure with criteria for determining a best-fitting prediction line.

✏ Interpret the Results in the Context of the Original Question

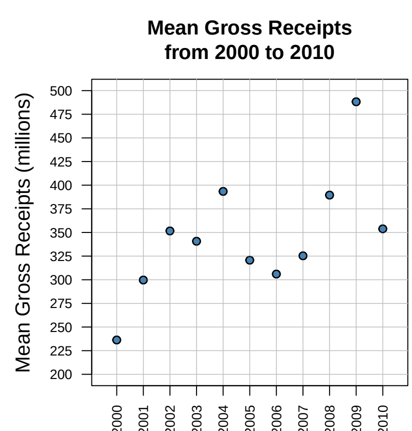
1. Place students into groups of three. Ask each group to discuss their answer to the original question, "By how much, if any, are the average gross receipts for movies increasing over time?"

   Each group should prepare a report on their findings. Your students' report should be based on the data collected, their scatterplots, and the rate of change. The report should include key calculations and graphs.
2. Ask your students if they think the trend they observed in the scatterplot will continue. Ask them to predict what the mean gross receipts for the top three movies will be in 2011. Ask them how close their prediction is to the actual mean. Why did they not match? **Note:** From Figure 5.4.3, a reasonable prediction would be $425 million. The actual results were Harry Potter and Deathly Hallows Part 2, $381.0; Transformers: Death of the Moor, $352.4; and The Twilight Saga: Breaking Dawn Part 1, $280.3. The mean of these top three grossing movies is $337.90.

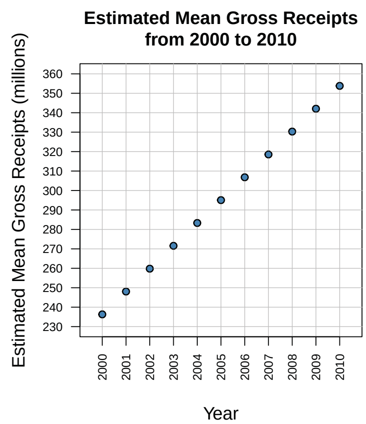
Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

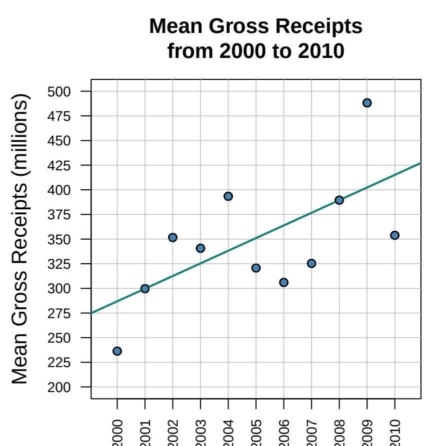
In our communications class last week, we were looking at old silent movies and comparing them to the high-tech ones of today. We were wondering how much movies make. We decided that a neat question to investigate in our mathematics class would be, "By how much, if any, are the average gross receipts for movies increasing over time?" From a website, we found a listing of gross receipts for movies year by year. We decided to look at the top three money-making films for the years 2000–2010 and then take the mean of the three to use as our data. The data are in millions of dollars by the way. To see if there was any relationship or trend, we drew a scatterplot, which is called a time series plot since time would be on the horizontal axis. Here is our plot:



We concluded from the graph that there is a positive relationship between time and mean gross receipts. That means, as we look at years from 2000 going up to 2010, mean gross receipts for those years generally increase. Of course, in some years, receipts went down, but overall there was an upward trend. To see the ups and downs, we calculated them and then found their mean. The average change in mean gross receipts was $11.75 million. In words, if the gross receipts changed a constant amount from year to year between 2000–2010, then that constant amount would be $11.75 million. So, we also looked at this by saying that if all our data points fell on a straight line exactly, then the slope of that line would be 11.75. Here's a graph of what that constant situation would look like:



But, of course, our real data did not fall on a straight line. So, as a final part of our analysis, we looked at our original data in its plot and drew a line through the data that we thought would fit the data pretty well. We went through the points and picked on (2001, 299.7) and (2008, 389.5). Here is the plot with our prediction line on it:



The slope for our estimated real data line was (389.5 – 299.7) / (2008 – 2001) = $12.83 million. It is higher than the average one of $11.75 million. It's kind of hard to say which method is right. The constant method averaged over the ups and downs, which smoothed things over. The picking two points method is very dependent on which points were chosen, but we think we did a good job because, looking at the graph, the line balances the points fairly well. So, we like our two-point method better.

To use our line to predict what mean gross receipts might be in 2011, we see from the graph that a prediction would be around $425 million. To be more correct, our slope is $12.83 increase per year. We know that (2008,389.5) lies on our line. Since 2011 is three years from 2008, our prediction for 2011 is 389.5 + 3*12.83 = $427.99 million.

We checked the website and found that the actual top movies in 2011 grossed a mean of $337.90, considerably less than our prediction. One reason is that the economy is not very good and people don't have as much money to spend on going out.

Assessment with Answers

A group of students was interested in answering the question, "By how much have the winning times of the past 15 Olympic Games men's 100-meter dash decreased? Table 5.4.6 shows the data they collected for the years 1952–2008.

Table 5.4.6 Winning Times for Men's Olympic 100-Meter Dash

Winning Times (seconds) - Olympic Games 100-Meter Dash – Men								
Year	1952	1956	1960	1964	1968	1972	1976	
Time	10.40	10.50	10.20	10.00	9.95	10.14	10.06	
Year	1980	1984	1988	1992	1996	2000	2004	2008
Time	10.25	9.99	9.92	9.96	9.84	9.87	9.85	9.69

Note that, in 1992, the original winner was Ben Johnson of Canada, who ran the dash in 9.79 s, but he was stripped of the medal after testing positive for steroid use. Figure 5.4.4 is a time series plot with the year on the horizontal axis and the dash time on the y-axis.

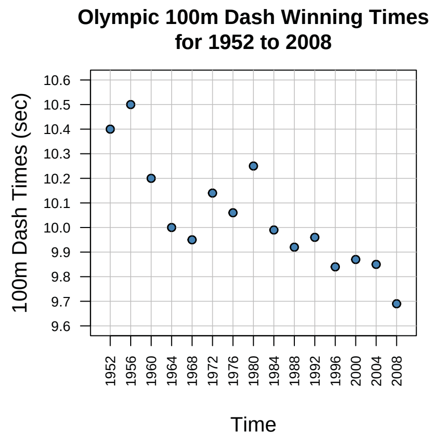


Figure 5.4.4 Time series plot of men's Olympic 100 m dash winning times for 1952–2008

Write a report starting with answering the question, "By how much have the winning times of the past 15 Olympic Games men's 100-meter dash decreased?" Include the following:

- A description of the trend you observe in the data. The trend in the winning times for the 100-meter dash has been decreasing.
- Identification of the years in which the Olympic 100-meter time was higher than the previous Olympic 100-meter time. In 1956 and in 1980, the winning times were higher than the previous Olympic 100-meter time.
- An appropriate graph with a line drawn through the points (1952, 10.4) and (2008, 9.69) and, by using these two points, the rate of change of the Olympic 100-meter times. The rate of change is -0.01. See Figure 5.4.5.
- A written explanation of what the rate of change of the times represents in the context of this investigation. A rate of change of -0.01 means that the winning time for the 100-meter dash decreases by about 0.01 seconds every year. Or, the rate of change of -0.01 means that every Olympics (four years), the winning time for the 100-meter dash decreases by about 0.04 seconds.

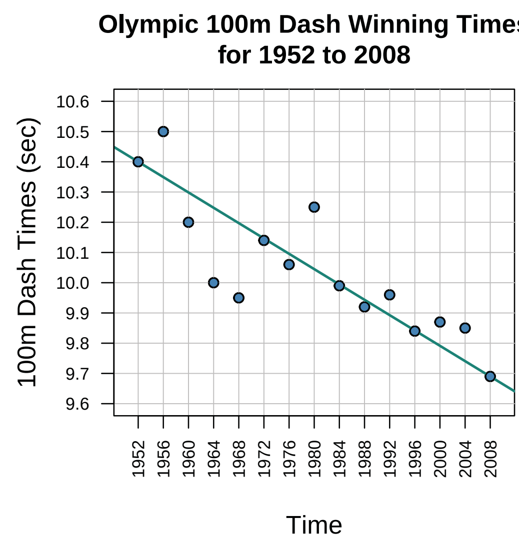


Figure 5.4.5 Time series plot of men's Olympic 100 m dash winning times for 1952–2008 with trend line

Extension

This investigation can be extended into developing the equation of a line. Develop the equation of the line your students drew through the points (2001, 299.7) and (2008, 389.5). They should find the rate of change (slope) and write the equation in point-slope form. Then, they should find the y-intercept and interpret this value. Students also could write the equation of the line in slope-intercept form and discuss the usefulness of the y-intercept in this problem.

References

Burrill, G., and P. Hopfensperger. 1998. *Exploring linear relations.* New York, NY: Dale Seymour.

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework.* Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics, *www.corestandards.org*.

IMDb, *www.imdb.com/boxoffice/alltimegross*.