Scatter It! (Using Census Results to Help Predict Melissa's Height)

Susan Haller
St. Cloud State University
skhaller@stcloudstate.edu

Published: March 2012

Overview of Lesson
In this lesson, students explore the relationship between age and height in order to help a hypothetical student predict her height. Students will examine data that will enable them to create a scatterplot and approximate a line of best fit. The scatterplot and line of best fit will be used to predict height. The slope of the line of best fit will be interpreted in context.

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 question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level A activity.

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.
5. Use appropriate tools strategically.

Common Core State Standards Grade Level Content (Grade 8)
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.
8. SP. 2. Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots 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 the 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 Standards for Grades 6-8
   Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them:
   - 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, box plots, and scatterplots.

Select and use appropriate statistical methods to analyze data:
- use observations about differences between two or more samples to make conjectures about the populations from which the samples were taken;
- discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots.

Develop and evaluate inferences and predictions that are based on data:
- make conjectures about possible relationships between two characteristics of a sample on the basis of scatterplots of the data and approximate lines of fit;
- use conjectures to formulate new questions and plan new studies to answer them.

Prerequisites
Students will have knowledge of and experience with graphing points on a coordinate system.
Students will have knowledge of and experience with constructing and evaluating properties of lines.
Students will have experience using technology to create graphs of data.
Students will have completed Scatter It! (Predict Billy's Height).

Learning Targets
Students will correctly construct a scatterplot, both manually and with the use of technology.
Students will analyze the scatterplot visually, looking for trends and patterns. In addition, students will approximate a line of best fit manually, and then create one via technology.

Time Required
One-half class period (assuming students have completed the previous Scatter It! lesson)

Materials Required
- Graphing calculators or computers with software capable of creating scatterplots.
- Spaghetti noodles.
- Graph paper.
- Internet access.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

I. Formulate a Question
Begin the lesson by formulating a hypothetical situation:

Melissa is a first-grade student who wants to know how tall she will be when she is in tenth-grade. Her class is studying the census, and Melissa wonders whether information from the census can help her determine her age when she is in tenth-grade.
Ask students to write some questions that they would be interested in when using the census to investigate the relationship between age and height. Some possible questions might include:
1. What kinds of information does the census have? Can anyone use this information?

2. How much information is available on the census? How do we know how much of the information we should use?
3. Is the relationship between age and height the same for male and female students? If so, how are they similar? If not, how are they different? Will we know whether the information comes from a male or female student?
4. Can we use the relationship between age and height from students we don't know in order to predict the height of an individual student?
5. Will we get the same results every time we take a sample from the census?

II. Design and Implement a Plan to Collect the Data
Melissa wants to use data from the census to help her predict her age when she is 15.

The table below shows data for 24* randomly selected females from the Census at School web site: http://www.amstat.org/censusatschool/index.cfm. Students recorded their age in years and height in centimeters.

Table 1. Age and height for 24 randomly selected female students.

Age	Height	Age	Height	Age	Height
12	153	10	170	12	152.5
11	150	11	144	13	156
16	160	16	160	14	156
14	166	14	171	17	165
12	153	15	171	12	156
14	162	14	151	15	170.6
15	164.5	11	154	11	157
13	152	10	147	11	142

*25 female students were selected, but one student did not have a height recorded, so her data was eliminated.

When selecting information to be included from the Census web site, note that not all states have information recorded. You may have to 'select all' or select a different state if the state you wish to sample does not have any data. For the purposes of this exercise, a sample of size 25 works very well. Also, because we wish to forecast the age of a female student, we limit the gender to female and the age to those from grades 4 – 10.

After selecting the information, delete all columns except the two that pertain to age and height. Note that some students may have recorded height in inches rather than in centimeters. In this case you may either choose to convert to cm or eliminate the data.

Use a spreadsheet or graphing calculator to graph the data.

III. Analyze the Data
After the scatterplot of the student data has been constructed, ask students to share anything they notice within the pattern of growth – does there seem to be a relationship between age and height? Does this relationship appear to be linear? Discuss the properties of a line of best fit,

Contact **Author** for permission to use materials from this STEW lesson in a publication

and have students place a piece of spaghetti in their scatterplot to approximate the line of best fit for the data. Next, have students locate the height of a female at age 15. Students should compare their results with the rest of the class (predictions will vary because the lines will be slightly different).

Students should then use the available technology to calculate the line of best fit for the data. Alternatively, the teacher can calculate it for the students. In this case, the line of best fit is approximately $y = 2.37x + 126.6$.

The graph in Figure 1 shows the relationship between age and height for the randomly sampled female students.

Figure 1. Scatterplot for age and height.

IV. Interpret the Results
The students should next consider the line of best fit in context of the scenario presented. Students should recognize that the slope indicates that, for the female students in this sample, the average height increases 2.37 cm for every year of age. Students can then use the equation for the line of best fit and calculate that the height of the average female student at age 15 should be between 162 and 163 cm.

Students should also assess the relationship between the graphed data and the line of best fit. For example, many data points are below the line of best fit, and many above. This means that the associated heights are below average and above average, respectively.

The teacher should engage the students in a discussion about the constraints for the line of best fit. For example, the equation would not be appropriate to use with adults.

Assessment

1. The table below shows the age (in years) and height (in cm) of 24 randomly selected female students.

| Age	Height	Age	Height	Age	Height	Age	Height
15	152	15	164	14	165	12	152.5
15	161	14	175	14	155	11	139
10	149	16	168	12	157.5	15	161
14	163	15	161	11	148	14	168
11	162	11	153	11	150	10	145
16	155	14	164.5	12	142.2	14	163

(a) Construct a graph of the data and describe the relationship between the female students' age and height.

(b) Find the equation for the line of best fit for the data.

(c) Predict the height of a randomly selected female at age 17.

Answers

(a)   As age increases, so does height. The relationship appears to be somewhat linear.

Age/Height Relationship
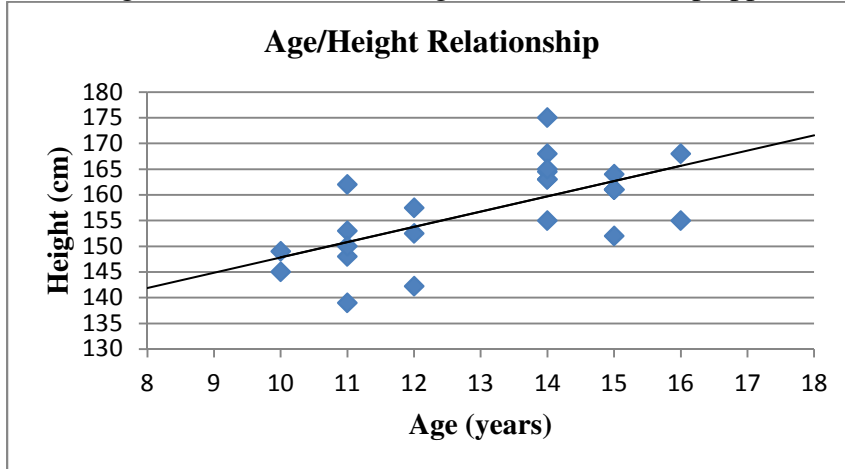
(b) The equation for the line of best fit is y = 2.97x+118.15.
(c) According to the line of best fit, the height should be between 168 and 169 cm.

Possible Extensions
Students can explore the Census at School data to find other variables that may be related. For example, age and reaction time, or foot length and height. Students might also select data from two different geographic regions and compare the age/height relationship, or make the same comparison by segregating by favorite sport.

References
1. This lesson was adapted from an investigation (Scatter It!) found at Census at School – New Zealand (http://www.censusatschool.org.nz/classroom-activities/).

2. Census data was obtained from Census at School (http://www.amstat.org/censusatschool/index.cfm).

Scatter It! Activity Sheet

Melissa is a first-grade student who wants to know how tall she will be when she is in tenth-grade. Her class is studying the census, and Melissa wonders whether information from the census can help her determine her age when she is in tenth-grade.

1. What are some questions you have about using the census to investigate the relationship between age and height?

2. The table below shows data for 24 randomly selected females from Census at School http://www.amstat.org/censusatschool/index.cfm. Students recorded their age in years and height in centimeters. The original sample included 25 students, but one student was eliminated because her height was not recorded.

Age	Height		Age	Height		Age	Height
12	153		10	170		12	152.5
11	150		11	144		13	156
16	160		16	160		14	156
14	166		14	171		17	165
12	153		15	171		12	156
14	162		14	151		15	170.6
15	164.5		11	154		11	157
13	152		10	147		13	165

Which variable, age or height, is the independent variable? Why?

3. Use a graphing calculator or computer spreadsheet to construct a graph showing the relationship between the students' ages and heights.

4. Describe the relationship you notice about age and height.

Contact Author for permission to use materials from this STEW lesson in a publication

5. Calculate the line of best fit for the data.

What does the slope of the line of best fit mean in this situation?

Use your line of best fit to approximate Melissa's height when she is 15.

6. Should all classmates get the same equation for the line of best fit for the data above? Why or why not?

7. Should all classmates get the same equation for the line of best fit if each collected their own random sample? Why or why not?