How Tall and How Many?

Subhash Bagui
University of West Florida
sbagui@uwf.edu

STatistics Education Web
STEW
online journal of K-12 statistics lesson plans

Published: March 2012

Overview of Lesson
In this activity, students will conduct an investigation to find out if there might be a relationship between a person's height and the number of siblings of that person. Two pieces of information will be collected for each student (subject). This form of data is called paired data. Students will examine the mean height in the class and the mean number of siblings in the class. Students will also construct a scatterplot. Conclusions will be drawn based on (statistical) calculations and an examination of the scatterplot.

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 B 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 for Mathematical Practice (Grade 8)
8. SP. 1. Construct and interpret scatter plots 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 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:

- discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots;

- find, use, and interpret measures of center and spread, including mean and interquartile range.

Develop and evaluate inferences and predictions that are based on data:

- use observations about differences between two or more samples to make conjectures about the populations from which the samples were taken.

Prerequisites

Students will have knowledge of making measurements (in inches) using a tape measure. Students will have knowledge of how to organize data in a table. Students will have knowledge of how to create boxplots and scatter plots of paired data. Students will have knowledge of how to interpret descriptive statistics.

Learning Targets

Students will be able to collect and record paired data, to construct boxplots for each variable, and to create a scatterplot for the data set using technology. They will also learn how to create and conduct a statistical investigation. In addition, students will be able to analyze and interpret the collected data.

Time Required

Approximately one 45-50 minute class period; some discussion time the following class period might be necessary.

Materials Required

Graphing calculator or computer with spreadsheet or statistical software that can be used to create and print boxplots and scatter plots, 6' measuring tape, data recording sheet, poster board to display the data.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

I. Formulate Question(s)

Begin the lesson by telling the class that it is important to examine relationships between two variables. For example we might want to know if the weight of a person depends closely on their height or age, or if the height of a person depends on their family size or their age. Ask students if they think there might be a relationship between a person's height and the number of people in that person's family. Is this relationship linear or non-linear? Discuss with the class what types of graphical displays might be used to illustrate relationships between variables. Discuss how the graph would look if there is a relationship between height and family size. Discuss how the graph would look if there is no relationship between height and family size.

Ask students to write some questions that they would be interested in investigating about students' height and number of siblings. Some possible questions might be:

1. What is the mean (representative) height of the class? What is the median height of the class? What is the shortest height in the class? What is the tallest height in the class?
2. What is the typical family size in the class? What is the largest family size? What is the smallest family size?

II. Design and Implement a Plan to Collect the Data

In the data collection phase ask students what kind of measurements should be made? Make sure that students talk about how to make the measurements accurately and with precision. Ask the class who should measure the students. More than one person might be helpful in the data collection phase. It might be a good idea for one person to measure all students and for another to record their height along with the number of siblings they have. Students should discuss how a student must stand at the time of measurement and whether or not shoes should be worn. Discuss if data should be collected on family size or on number of siblings. Would it make any difference? Also define the term sibling (such as brother/sister, half brother/half sister etc.). This way a protocol for measurements can be developed and it will ensure consistency from measurement to measurement.

Measure all students and record their heights (in inches) and number of siblings in a data table. A sample class data set is shown in the table below; a blank data table is provided on the Activity Sheet on page 10.

Table 1. Sample class data.

Student Name	Height (inches)	# of Siblings
Adam	64	4
Mallory	67	3
Marianne	65	7
Patrick	64	1
Darryl	67	0
Taylor	63	7
Tasha	62	4
Wes	69	8
Will	71	2
Amanda	67	1
Dave	68	2
Jason	62	2
Jake	64	0
Darcy	70	3
Marissa	68	1
Paul	72	5
Teal	67	3
Zac	60	2
Ian	64	4
Alan	67	1
Ambreia	62	3
Davis	69	0
Greg	64	1
Javin	67	1

III. Analyze the Data

Different statistical tools are used for analysis of different questions. For example, one may just want to analyze the heights or the number of siblings separately. Also one may want to analyze the data by taking into account both height and number of siblings jointly to determine if there might be a relationship between them.

To analyze the heights and the number of siblings separately the class can calculate measures of center and spread and create a boxplot. To create a boxplot one needs the 5-number summary: minimum, first quartile (25^{th} percentile, Q_1), second quartile (median, 50^{th} percentile, Q_2), third quartile (75^{th} percentile, Q_3), and maximum. First these five numbers are plotted on a line extended from the minimum to the maximum and then a box is created around Q_1 and Q_3 with lines drawn at the first quartile, the second quartile, and the third quartile. The difference between the third and first quartiles is called the interquartile range (IQR).

Descriptive Statistics for the 24 sample class heights are calculated. The numerical calculations show that the mean height of the class is 65.96 inches. The median height is 67.00 inches. The teacher can discuss with the class that the median represents the 50^{th} percentile of the distribution of the class heights. About half of the heights are less than 67.00 inches and another half of the heights are more than 67.00 inches. The shortest height in the class is 60.00 inches; the tallest height is 72.00 inches. About one fourth (25%) of the heights are below 64.00 inches (the first quartile) and one fourth (25%) of the heights are above 68.00 inches (the third quartile). About half (50%) of the heights are between 64.00 and 68.00 inches. The standard deviation of 3.11 inches provides a typical difference between the student heights and the mean.

Similarly, descriptive statistics for the number of siblings are given by: Mean=2.71, Standard Deviation=2.26, Min=0.00, First Quartile =1.00, Median =2.00, Third Quartile=4.00, and Maximum=8.00. These descriptive statistic values can be interpreted similar to above.

A boxplot for the class heights is shown in Figure 1. The boxplot depicts the 5-number summary of the class heights. The plot shows that the median height of the class is 67 inches. The middle 50% of the class heights ranges from 64 inches to 68 inches (as seen by IQR). Thus, here IQR = 68-64 = 4 inches. The shortest height is about 60 inches and the tallest height is about 72 inches. Similarly, a boxplot can be created for the number of siblings.

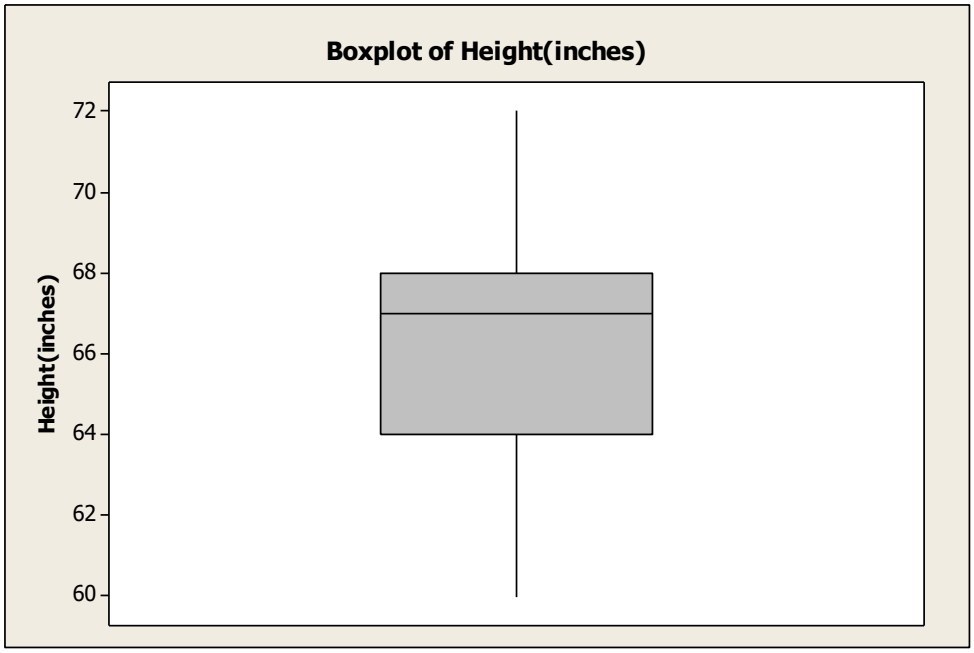


Figure 1. Boxplot for sample class heights.

In order to examine whether there might be a relationship between height and family size have students use appropriate technology (graphing calculator, Excel, statistical software) to create a scatterplot of the height and number of siblings data. A scatterplot for the sample data is shown in Figure 2 below.

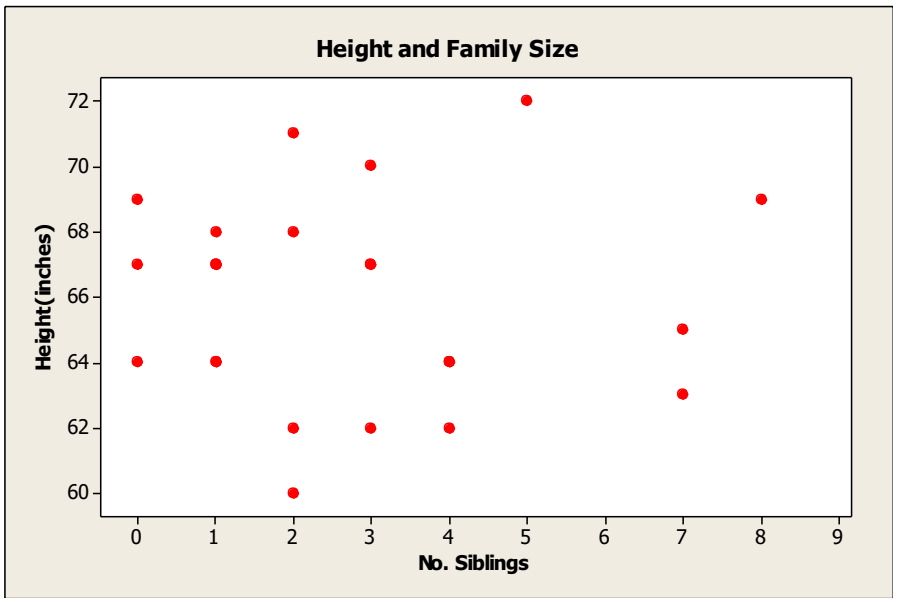


Figure 2. Scatterplot of heights and number of siblings.

Have students view the scatterplot carefully. Ask the class if there appears to be any evidence of a relationship between height and number of siblings. Ask students if the plot shows any linear or non-linear trend or no pattern. Discuss in the class that since the plot shows no pattern it is not possible to draw a best fitted line through the data.

IV. Interpretation of Results

The above scatterplot (Figure 2) shows no pattern (linear or non-linear) at all. In fact the plot shows a random pattern. Thus students may conclude that there is no relationship between heights of persons and number of siblings of those persons. So number of siblings of a person cannot predict the height of that person. Ask the class if this is result surprising? Tell the class that this result is not surprising, because number of siblings of a person cannot predict the height of that person. Ask the class to provide examples of variables that might be good predictors of height. Perhaps age might be a good predictor for height.

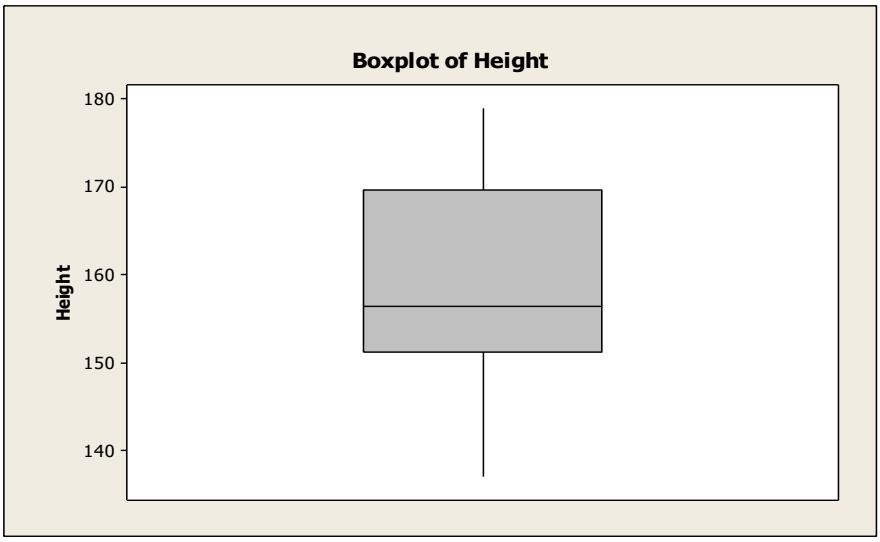
Assessment

1. A random sample of 20 students is selected and their heights and the number of people who lived in the person's household are recorded. The data are given in the following table:

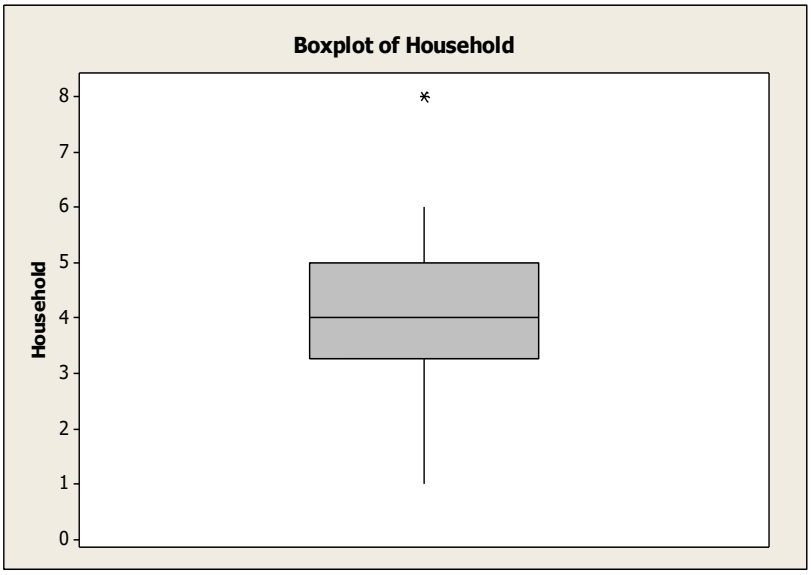
Student	Height (in cm)	# in Household
1	165	6
2	173	5
3	141	2
4	160	3
5	153	5
6	175	4
7	137	3
8	173	4
9	170	4
10	152	8
11	169	3
12	160	5
13	148	4
14	153	5
15	152	4
16	157	1
17	156	5
18	151	5
19	179	5
20	147	4

(a) Calculate the 5-number summary to create a boxplot for the heights of students.
(b) Construct the boxplot for heights.
(c) What is the mean height? What is the median height? What are the shortest and tallest heights? Give two numbers that cover the middle 50% of the distribution of the heights.
(d) Construct the boxplot for number in household.
(e) What is the typical household size? What is the median number in a household? What is the smallest and largest number in a household? Give two numbers that cover the middle 50% of the distribution of the number in a household.
(f) Using appropriate statistical software create a scatter plot of height vs number in household.
(g) Does there appear to be any evidence of relationship between height and number in household? How would you describe the relationship (linear or non-linear or no relationship)?

Answers:
(a) 5-number summary: Minimum=137.00, First Quartile=151.50, Median=156.50, Third Quartile=169.50, and Maximum=179.00.
(b) See the boxplot for heights below:

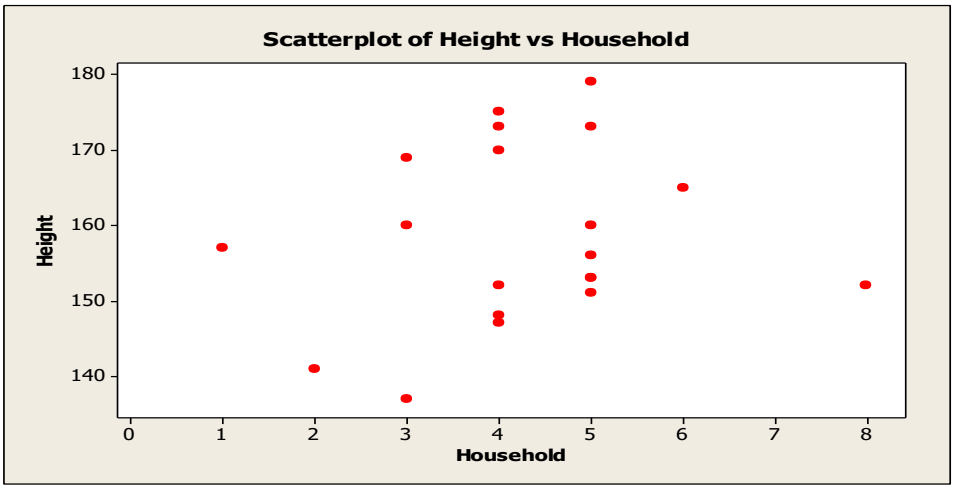


(c) Mean height=158.55, Median height=156.50, Shortest height=137.00, Tallest height=179.00, The middle 50% of the distribution of the heights are between 151.50 cm and 169.50 cm.
(d) See the boxplot for number in households below:



- In the above plot * shows an outlier (extreme value) in the household sizes. Note that there is a household of size 8 in the sample.

(e) Typical (mean) household size = 4.25, Median=4.00, Smallest household size=1.00, Largest household size=8.00. The middle 50% of the distribution of the household sizes are between 3.50 and 5.00.
(f) The scatter plot of height vs number in household is given below:



(g) The scatter plot does not show any linear or non-linear pattern. In fact it shows no pattern. Since the scatter plot shows a random pattern we cannot conclude that there exists a relationship between height of a person and the number of persons in the household.

Possible Extensions

Intuition tells us that age can be a good predictor for height. So an extension of this experiment is to use age instead of number of siblings. Start by asking the class that if they expect the relationship between height and age to be linear.

References

1. Adapted from an activity created by Paul J. Fields, Ph.D. for the *American Statistical Association Meeting Within a Meeting Program for Middle School Teachers* (2008).

2. *You and Michael* by Stephen Miller 2012, http://www.amstat.org/education/STEW/.

3. *Sampling in Archaeology* by Mary Richardson 2012, http://www.amstat.org/education/STEW/.

How Tall and How Many Activity Sheet

In this activity, you will collect data to assess if there might be a relationship between a person's height and the number of siblings of that person.

1. Describe the data collection process that will be used to answer the research question.

2. Record the height (in inches) and number of siblings for you and your classmates in the following table:

Student name	Height	# of Siblings

3. Sort the heights of the students from smallest to largest.

4. Using the sorted data, find the values needed to create a boxplot of the data.

5. Construct the boxplot for heights:

6. What is the mean height of the class? What is the median height of the class? What is the shortest height in the class? What is the tallest height in the class? Give two numbers that cover the middle 50% of the distribution of the heights.

7. Sort the number of siblings from smallest to largest.

8. Using the sorted data, find the values needed to create a boxplot of the data.

9. Construct the boxplot for number of siblings:

10. What is the typical family size in the class? What is the median of the number of siblings in the class? What is the largest of number of siblings in the class? What is the smallest number of siblings in the class? Give two numbers that cover the middle 50% of the distribution of the number of siblings.

11. Do you think there is a relationship between students' height and number of siblings?

12. What type of graphical display might be used to exhibit the relationship? How would the graph look if there is a relationship between height and number of siblings? How would the graph look if there is no relationship between height and number of siblings?

13. Using appropriate statistical software create a scatter plot of height vs number of siblings.

14. Does there appear to be any evidence of a relationship between height and number of siblings? How would you describe the relationship (linear or non-linear or no relationship)?