

Bear Hugs

Subhash Bagui
University of West Florida
sbagui@uwf.edu

Mary Richardson
Grand Valley State University
richamar@gvsu.edu



Published: July 2012

Overview of Lesson

To give an awesome bear hug one needs really long arms, i.e., a long arm span. In this activity, students will study arm span lengths of middle school students using a sample of data collected from a data base called Census at School. Students may prefer to use data collected from their own class. To analyze the data students will compute measures of center (mean, median), quartiles (first, second, and third quartiles) and spread (range, Interquartile Range (IQR), standard deviation) for boys and girls combined and also separately. Students will also construct a comparative boxplot to compare arm spans of boys and girls. A scatter plot will be used to see if arm span of students gets larger as age increases. Conclusions will be drawn based upon analysis of the data and examinations of graphs in the context of questions asked about lengths of arm spans.

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 (Grades 7 and 8)

7. SP. 3. Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.
7. SP. 4. Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.
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.

Prerequisites

Students should have prior knowledge of making measurements (in cm) using a ruler, organizing data in a table, calculating descriptive statistics, creating boxplots, and how to interpret descriptive statistics.

Learning Targets

Students will be able to construct a comparative boxplot that compares two groups and a separate boxplot for combined data, and also to create a scatterplot for the data set using technology. In addition, students will be able to analyze, compare, and interpret the (statistical) analysis of the data.

Time Required

Approximately one 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, rulers with centimeters, data recording sheet.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

I. Formulate Question(s)

Begin the lesson by telling the class that to give an awesome bear hug you need really long arms so that your arms get all the way around the person you are hugging. In fact, not all arm span lengths of students are the same. So it is of interest to examine student's arm spans. More specifically, a statistical study can be conducted to study arm spans of students in the class.

Ask students to write some questions that they would be interested in investigating about students' arm span. Some possible questions might be:

1. What is the mean (representative) arm span of the class? What is the median (typical) arm span of the class? What is the shortest arm span in the class? What is the longest arm span in the class?
2. Are there differences in arm spans for boys and girls? If so, what are the differences?
3. Are arm span lengths related to age?
4. Are arm spans related to any other variables?

II. Design and Implement a Plan to Collect the Data

If students plan to use in class data, 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' arm span. 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 data, e.g., record students' arm spans, gender, and age. Before collecting data ask students to discuss whether to measure arm spans in inches or in centimeters. This way a protocol for measurements can be developed and it will ensure consistency from measurement to measurement. Measure all students in the class and record their arm spans, gender, and age.

In this activity, a random sample of 24 students is selected from the Census At School website, (<http://amstat.org/censusatschool/>), along with their arm span, gender, and age; this data is given in Table 1 below. A blank data table is provided on the Activity Sheet on page 12.

Table 1. Sample data from the Census At School website.

| Student | Arm Span (cm) | Gender | Age (year) |
|---------|---------------|--------|------------|
| 1 | 163 | M | 12 |
| 2 | 155 | F | 14 |
| 3 | 155 | F | 12 |
| 4 | 144 | M | 10 |
| 5 | 164 | F | 14 |
| 6 | 144 | M | 9 |
| 7 | 165 | F | 13 |
| 8 | 118 | F | 14 |
| 9 | 162 | F | 14 |
| 10 | 138 | F | 10 |
| 11 | 144 | M | 11 |
| 12 | 140 | F | 9 |
| 13 | 128 | M | 9 |
| 14 | 163 | M | 13 |
| 15 | 147 | F | 13 |
| 16 | 125 | M | 11 |
| 17 | 154 | M | 13 |
| 18 | 159 | F | 12 |
| 19 | 156 | F | 15 |

| | | | |
|----|-----|---|----|
| 20 | 175 | M | 13 |
| 21 | 175 | F | 15 |
| 22 | 130 | F | 9 |
| 23 | 182 | M | 14 |
| 24 | 150 | F | 12 |

Ask the students to explain why this is an observational study and not an experimental study. Tell students that data values are randomly sampled from an existing database where data were recorded from direct observation and measurements. In the existing (original) dataset nothing has been done deliberately to the students in order to record/collect data.

III. Analyze the Data

Different statistical tools are used for analysis of different questions. For example, the class can calculate the mean and spread from the data collected on arm span. A boxplot can be constructed from the same data set. The same analyses can be repeated separately for the boys and girls in class. Ask students to suggest graphs that might be useful to compare the arm span length data distributions for boys and girls. Comparative boxplots are appropriate for displaying this data. Ask students to suggest a plot that will help one to examine if there is a relationship between the length of student's arm and their age. The scatterplot is a useful graph to analyze a relationship between variables.

To analyze the arm spans, the class can calculate measures of center and spread and create boxplots. To create a boxplot one needs the 5-number summary: minimum, first quartile (25th percentile, Q_1), second quartile (median, 50th percentile, Q_2), third quartile (75th 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). That is, $IQR = Q_3 - Q_1$.

Descriptive Statistics for the 24 sample arm spans are calculated. The numerical calculations show that the mean arm span of the sample is 151.5 cm. The median arm span is 154.5 cm. The teacher can discuss with the class that the median represents the 50th percentile of the distribution of the arm spans. About half of the students in the sample have arm spans less than 154.5 cm and another half of the students in the sample have arm spans higher than 154.5 cm. The shortest arm span in the sample is 118.0 cm; the longest arm span is 182.0 cm. About one fourth (25%) of the students in the sample have arm spans below 142.0 cm (the first quartile) and one fourth (25%) of the students in the sample have arm spans above 163.0 cm (the third quartile). About half (50%) of the students in the sample have arm spans between 142.0 and 163.0 cm. The standard deviation (s.d.) of 16.42 cm provides a typical difference between the student arm spans and the mean.

The following rule can be used to find an outlier (extreme value) in a data set. An arm span (X) is called an outlier if $X > Q_3 + 1.5(IQR)$ or $X < Q_1 - 1.5(IQR)$. For the combined (boys and girls) data set $Q_1 = 142.0$, $Q_3 = 163.0$, and $IQR = Q_3 - Q_1 = 21$. So $1.5(IQR) = (1.5)(21.0) = 31.5$. Thus, $Q_3 + 1.5(IQR) = 163.0 + 31.5 = 194.5$ and $Q_1 - 1.5(IQR) = 142.0 - 31.5 = 110.5$. Any arm span value

(X) in the combined data greater than 194.5 cm or smaller than 110.5 cm would be an outlier. Similar rules can be formulated for boys and girls separately to find outliers in the respective groups.

The descriptive statistics for the sample boys' data are given by: $n=10$, Mean=152.2, Standard Deviation=18.77, Min=125.0, First Quartile =144.0, Median =149.0, Third Quartile=163.0, and Maximum=182.0. These descriptive statistic values can be interpreted similar to above.

Similarly, the descriptive statistics for the sample girls' data are given by: $n=14$, Mean=151.0, Standard Deviation=15.24, Min=118.0, First Quartile =140.0, Median =155.0, Third Quartile=162.0, and Maximum=175.0. These descriptive statistic values can also be interpreted similar to above.

A boxplot of arm spans for the combined sample is shown in Figure 1. The boxplot depicts the 5-number summary of the class heights. The plot shows that the median arm span for boys and girls is 151.5 cm. The middle 50% of the arm spans ranges from 142.0 cm to 163.0 cm (as seen by IQR). Thus here $IQR = 163.0 - 142.0 = 21.0$ cm. The shortest arm span is about 118.0 cm and the longest arm span is about 182.0 cm.

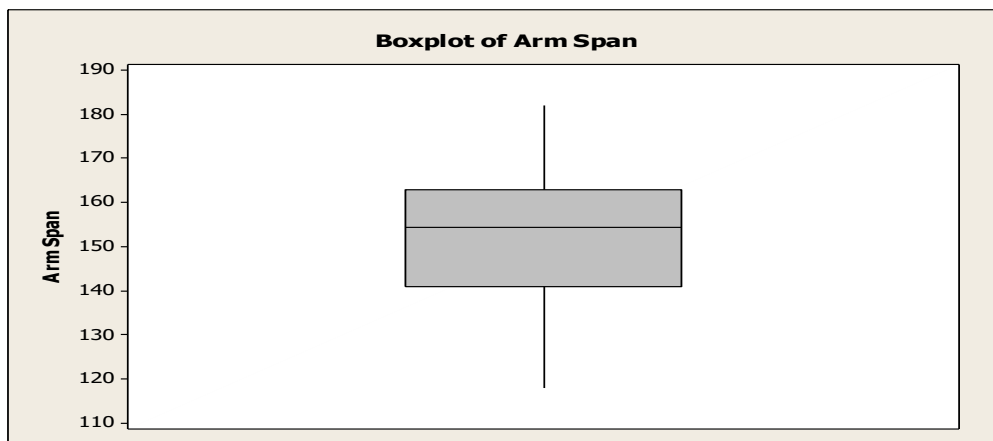


Figure 1. Boxplot for arm spans.

Similarly, a comparative boxplot can be created for boys and girls. Students may use appropriate technology (graphing calculator, Excel, statistical software) to create these plots. The comparative boxplot is shown in Figure 2 below.

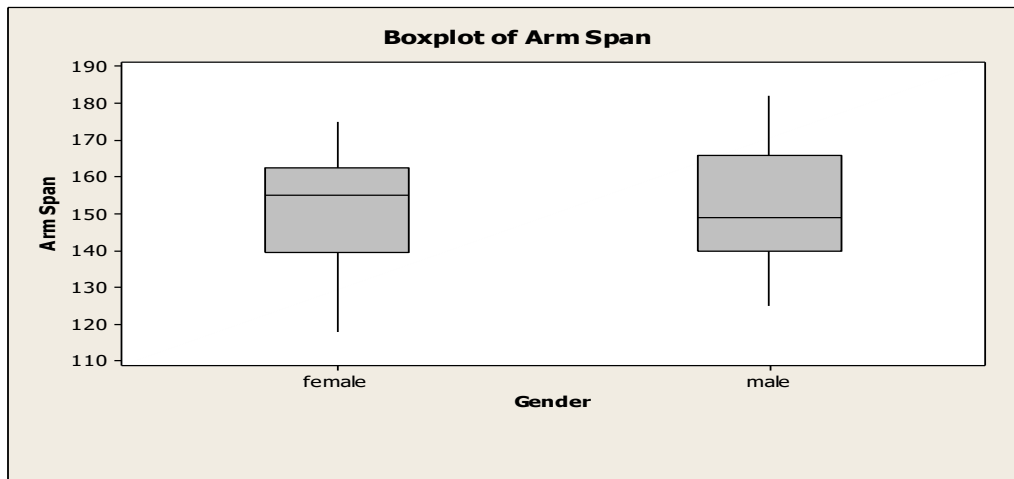


Figure 2. Comparative boxplot for boys and girls.

Have students view the above comparative boxplot for boys and girls. Ask the class if there appears to be any evidence of boys or girls having a higher median arm span length. Ask the class if the genders show similar variability in span. There are ten boys and fourteen girls in the sample. Different group sizes do not matter; but large samples would provide a better understanding of the underlying population. However, similar results would be expected if a new random sample of size 24 is selected.

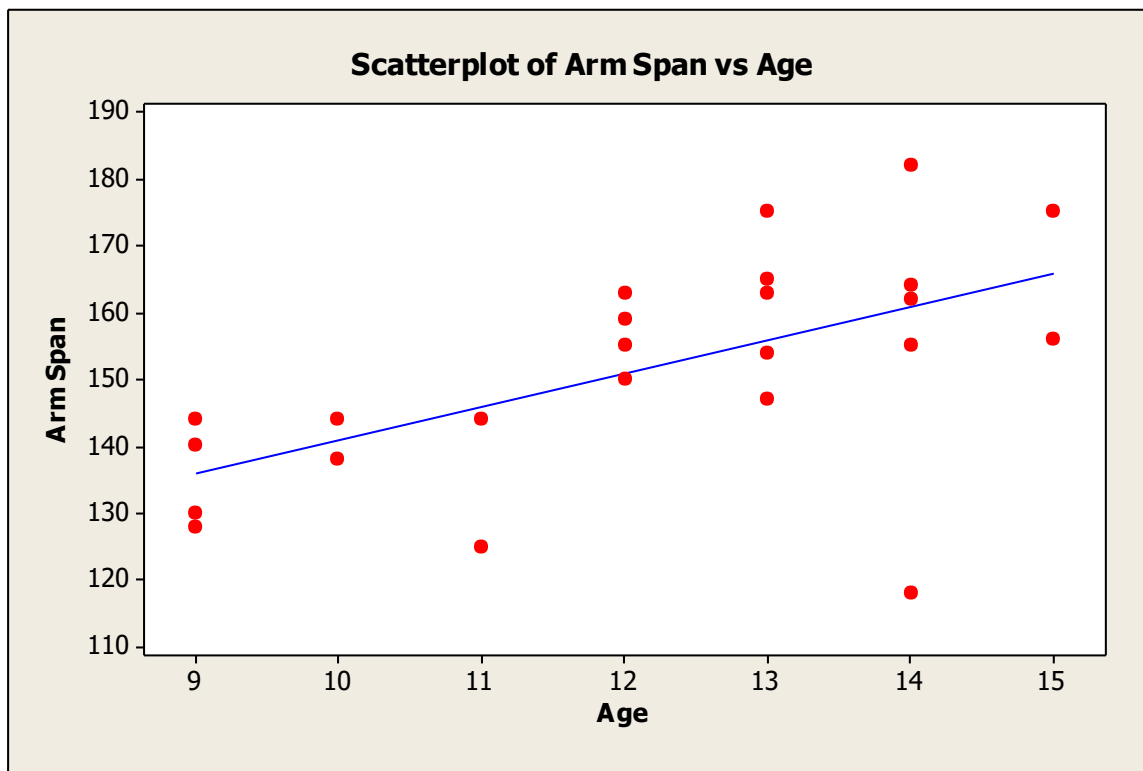


Figure 3. Scatter plot of arm span vs age.

In order to examine whether there might be a relationship between lengths of arm spans and age have students use appropriate technology (graphing calculator, Excel, statistical software) to

create a scatterplot of the arm spans versus age. A scatterplot of the arm spans versus age is shown in Figure 3 above.

Have students view the scatterplot carefully. Ask the class if there appears to be any evidence of an association between lengths of arm spans and age. Ask students if the plot shows any upward or downward trend. Discuss in the class that since the scatterplot shows an upward trend it is possible to draw a best fit line through the data.

IV. Interpret the Results

From Figure 1 we notice that the minimum value (118.0), first quartile (142.0), third quartile (163.0), and maximum value (182.0) are not symmetrically located around median (154.5). This indicates that the combined arm spans may not be symmetrically distributed. Also, the average arm span is 151.5 cm with a s.d. of 16.42 cm. The median is 154.5. The mean is less than the median because the distribution is slightly skewed to the left. Arm spans can be compared with a threshold value. This value can be chosen with an aim. Here we have chosen this value as 160 cm, a target circumference of a bear. Since the third quartile is 163, a little more than 25% of the students in the sample will have arms long enough to give a bear with a circumference of 160 cm a hug. Eight out of the twenty four students have arm spans longer than 160 cm. Thus approximately $(8/24)100\% \approx 33\%$ of the students will be able to give the bear a bear hug.

From the data analysis it appears that the mean arm span lengths for boys and girls are similar. Also from the comparative boxplot (Figure 2) we notice that girls tend to have higher median arm span length than the boys median arm span length. The minimum value, first quartile, third quartile, and the maximum value of arm span lengths are higher for boys than the girls. The upper half of girls' arm span lengths are less spread than the upper half of boys' arm span lengths. But they are in the opposite direction for the lower halves.

Next ask students to characterize the spread of the arm span distributions. The spread is generally measured using range (max-min), standard deviation (s.d.), or IQR. The spread summaries for boys and girls are shown in the table below:

| | Range | S.D. | IQR |
|-------|-------|-------|-----|
| Boys | 57 | 18.77 | 19 |
| Girls | 57 | 15.24 | 22 |
| | | | |

The spread for boys and girls with respect to range is the same. But otherwise boys' arm spans are more variable than girls' arm spans. Ask students to determine if there are any outlying values. Ask students if arm span lengths are related to any other variables?

Based on the scatter plot one may conclude that there seems to be a positive association (relationship) between a person's arm span length and his/her age. This makes perfect sense as students get older they are likely to get taller and bigger so one would expect their arm span to get longer. Also, for each age there can be a range of arm span lengths. For example, the arm span of thirteen year olds ranges from 147 to 175 cm. There is one student who is 14 and has an arm span of 118 cm. This is an unusual data point from rest of the data; may be considered as an outlier.

Assessment

1. A random sample of 30 students is selected from the Census At School website (<http://www.amstat.org/censusatschool/>) along with their arm span length, gender, and age. The data are given in the following table.

Table 2. Sample data from Census At School

| Student | Arm Span (cm) | Gender | Age |
|---------|---------------|--------|-----|
| 1 | 157 | M | 12 |
| 2 | 158 | M | 12 |
| 3 | 165 | M | 13 |
| 4 | 165 | F | 14 |
| 5 | 156 | F | 14 |
| 6 | 168 | F | 14 |
| 7 | 148 | F | 13 |
| 8 | 145 | F | 12 |
| 9 | 172 | F | 12 |
| 10 | 182 | F | 14 |
| 11 | 178 | M | 16 |
| 12 | 153 | M | 11 |
| 13 | 125 | M | 13 |
| 14 | 161 | F | 14 |
| 15 | 135 | M | 12 |
| 16 | 172 | F | 12 |
| 17 | 148 | F | 11 |
| 18 | 141 | M | 11 |
| 19 | 165 | M | 12 |
| 20 | 161 | M | 11 |
| 21 | 159 | F | 12 |
| 22 | 150 | M | 13 |
| 23 | 154 | F | 11 |
| 24 | 159 | M | 12 |
| 25 | 168 | M | 12 |
| 26 | 175 | M | 14 |
| 27 | 148 | M | 12 |
| 28 | 170 | F | 13 |
| 29 | 185 | M | 14 |
| 30 | 180 | M | 14 |

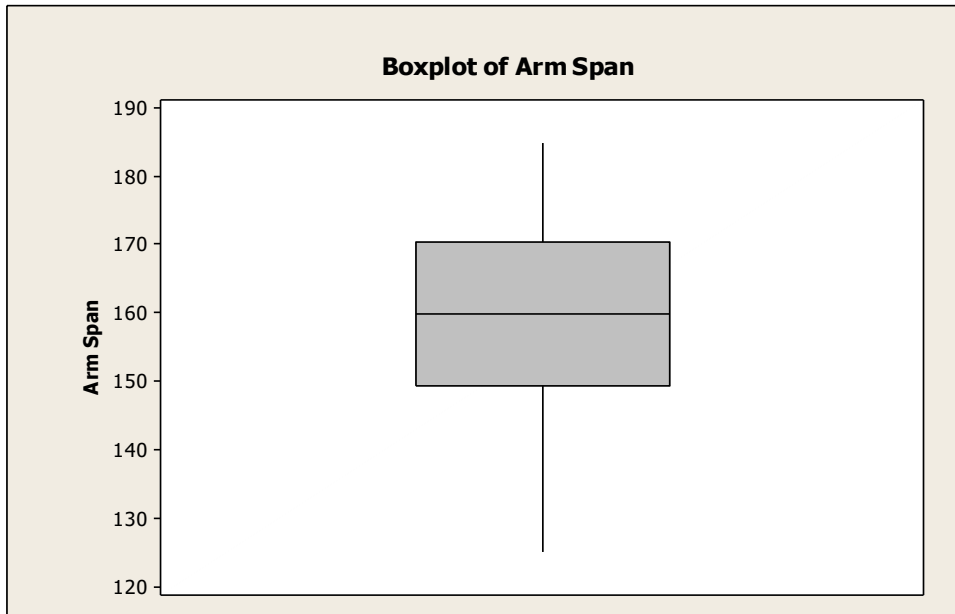
Using the above data answer the following questions:

- Calculate the 5-number summary to create a boxplot for the arm span lengths of students.
- Construct a boxplot for arm span lengths.

- (c) What is the mean arm span length? What is the median arm span length? What are the shortest and longest arm span lengths? Give two numbers that cover the middle 50% of the distribution of the arm spans. What is the range of the most common arm spans?
- (d) What is the mean arm span length for boys? What is the median arm span length for boys? What are the shortest and longest arm span lengths for boys? Give two numbers that cover the middle 50% of the distribution of the arm span lengths of the boys. What is the range of the most common arm span lengths for boys?
- (e) What is the mean arm span for girls? What is the typical arm span for girls? What are the shortest and longest arm span lengths for girls? Give two numbers that cover the middle 50% of the distribution of the arm span lengths of the girls. What is the range of the most common arm span lengths for girls?
- (f) Using appropriate technology (graphing calculator, Excel, statistical software) construct a comparative boxplot. Do boys or girls generally have bigger arm spans? Are there any outliers for boys or girls?
- (g) Using appropriate technology (graphing calculator, Excel, statistical software) create a scatter plot of age vs arm span.
- (h) Does there appear to be any evidence of an association (relationship) between age and arm span length? How would you describe the relationship (positive association, linear or non-linear or no relationship).

Answers

- (a) 5-number summary: minimum=125.0, first quartile=150.0, median=160.0, third quartile=170.0, and maximum=185.0.
(b) See the boxplot for arm span lengths below:

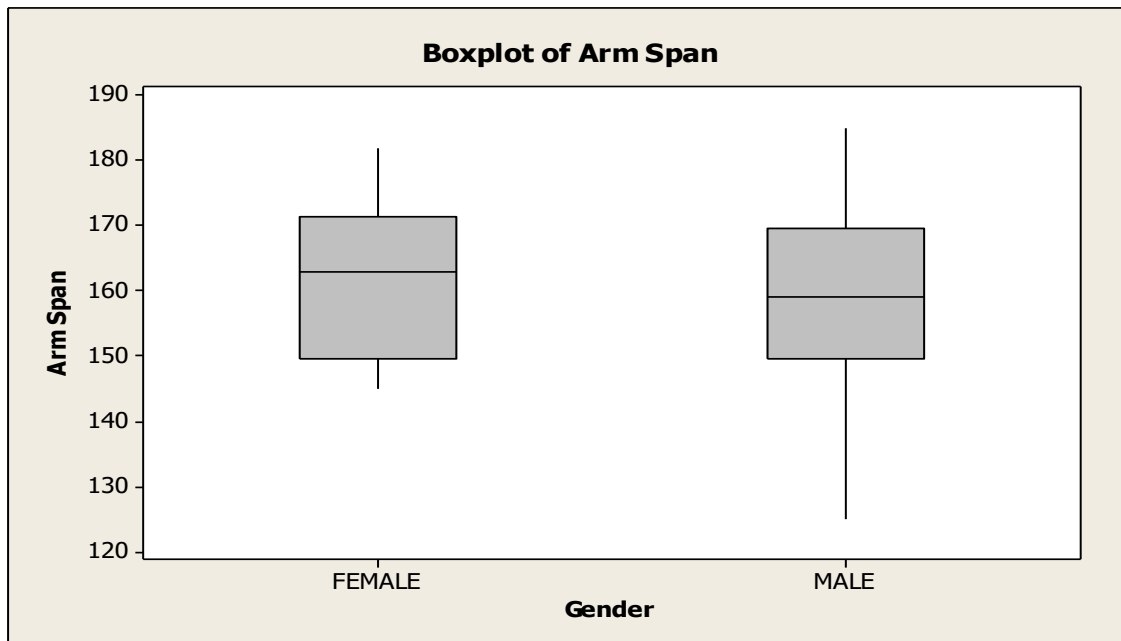


(c) Mean arm span= 160.1, median arm span=160.00, shortest arm span=125.0, longest arm span=185.0. The middle 50% of the distribution of the arm span lengths are between 150.0 cm and 170.0 cm. Most common arm span lengths range from 150.0 to 170.0 cm.

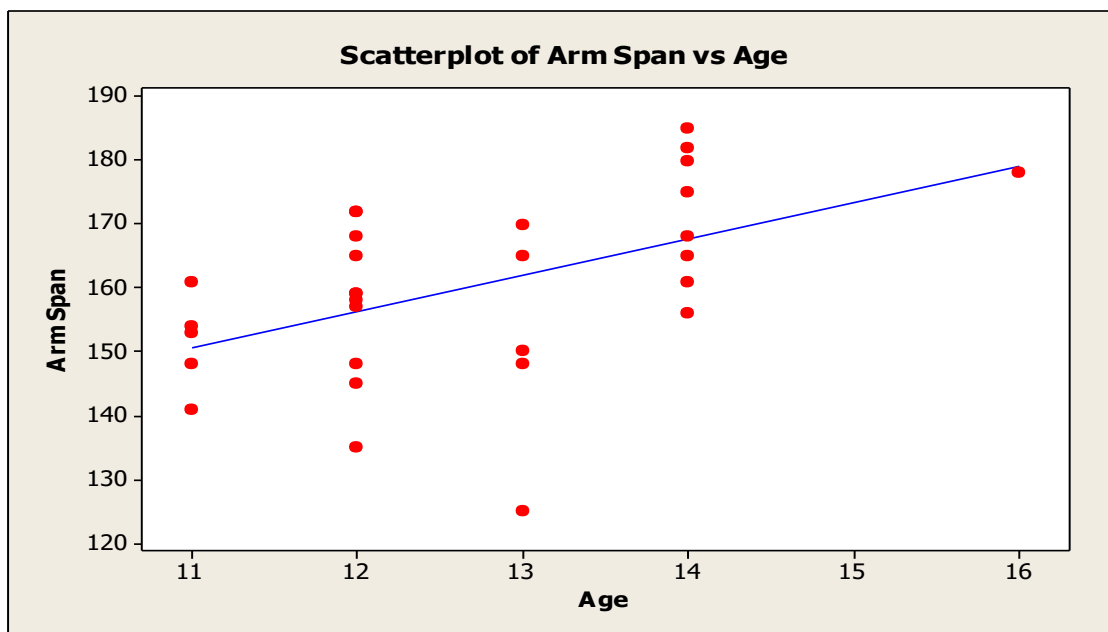
(d) Mean arm span length for boys=159.0, median arm span length for boys=159.0, shortest arm span length for boys=125.0, longest arm span for boys=185.0. The middle 50% of the distribution of the arm span lengths for boys are between 149.0 cm and 171.5 cm. Most common arm span lengths for boys range from 149.0 to 171.5 cm.

(e) Mean arm span length for girls= 161.54, median arm span length for girls =161.0, shortest arm span length for girls =145.0, longest arm span length for girls=182.00. The middle 50% of the distribution of the arm span lengths for girls are between 151.0 cm and 171.0 cm. Most common arm span lengths for girls range from 151.0 to 171.0 cm.

(f) See the comparative boxplot below. The median arm span length for females is higher than the median arm span length for males. There are no outliers in either of the two groups.



(g) See the scatter plot age vs arm span below:



(h) Based on the scatter plot one may conclude that there seems to be a positive association (relationship) between a person's arm span length and his/her age. The association (relationship) has a positive linear trend. But note that the arm span of twelve year olds ranges from 135 to 172 cm. There is one student who is 13 has an arm span of 125 cm. This is an unusual data point from rest of the data; and may be considered as an outlier.

Possible Extensions

Intuition tells us that height can be a good predictor of arm span length. So an extension of this activity is to see if there is a linear relationship between arm span length and height of students. Start by asking the class if they expect the relationship between height and arm span length to be linear. One can properly interpret and infer about the estimated slope/intercept parameters by fitting a regression line. Also, this activity can be extended by simply doing a two sample t-test. It can be determined if there is a statistically significant difference between the mean arm span lengths of two groups, boys and girls.

References

1. This activity is based on the activity by Joanne Woodward, titled *Bears Hugs 1-2-3* on the New Zealand Census at School website <http://www.censusatschool.org.nz/>.
2. *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) Report, ASA, Franklin et al., ASA, 2007, <http://www.amstat.org/education/gaise/>.
3. Census At School website, May, 2012 <http://www.amstat.org/censuschool/>.

11. What is the mean arm span for boys? What is the median arm span for boys? What are the shortest and longest arm spans for boys? Give two numbers that cover the middle 50% of the distribution of the arm spans of the boys. What is the range of the most common arm spans for boys?

12. What is the mean arm span for girls? What is the median arm span for girls? What are the shortest and longest arm spans for girls? Give two numbers that cover the middle 50% of the distribution of the arm spans of the girls. What is the range of the most common arm spans for girls?

13. Using appropriate technology (graphing calculator, Excel, statistical software) or by hand construct a comparative boxplot. Do boys or girls generally have bigger arm spans? Are there any outliers for the boys or girls?

14. Using appropriate technology (graphing calculator, Excel, statistical software), create a scatter plot of arm span length vs age.

15. Does there appear to be any evidence of association (relationship) between age and arm span length? How would you describe the relationship (positive association, linear or non-linear or no relationship).