Walk the Line

Maryann Huey
Drake University
Maryann.Huey@drake.edu

Published: August 2012

Overview of Lesson
In this activity, students will conduct an investigation to collect data to determine how far students can walk blindfolded down the length of a hallway before they deviate from a straight walking path and go 'out of bounds'. This data will be used to test the theory that humans are not able to walk in a straight line without being able to see landmarks. Students will test the data set for normality and build a 95% confidence interval to estimate the mean distance students can walk.

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 C 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.

Common Core State Standards Grade Level Content (High School)
S-ID. 1. Represent data with plots on the real number line (dot plots, histograms, and box plots).
S-ID. 4. Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.

NCTM Principles and Standards for School Mathematics
Data Analysis and Probability Standards for Grades 9-12
    Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them:
- understand the differences among various kinds of studies and which types of inferences can legitimately be drawn from each;
- compute basic statistics and understand the distinction between a statistic and a parameter.

    Select and use appropriate statistical methods to analyze data:
- for univariate measurement data, be able to display the distribution, describe its

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

shape, and select and calculate summary statistics.

Develop and evaluate inferences and predictions that are based on data:
- understand how sample statistics reflect the values of population parameters and use sampling distributions as the basis for informal inference.

Prerequisites
- Students should be able to calculate the mean of a data set, determine the median, and find the standard deviation.
- Students should be able to construct a box-plot from a data set.

Learning Targets
Students will learn how to test a data set for normality, and construct an associated confidence interval about the mean. Students will learn that assumptions about the underlying population characteristics are important to consider when generalizing from a sample and that not all data sets can be modeled as a normal distribution. Students will also discuss what generalizations can be inferred from a sample.

Time Required
One block period or two 50-minute class periods.

Material Required
1. A 100-yard, straight, and flat surface roughly 5 feet wide or narrower, marked by increments of 5 feet in length. A long school hall way is ideal.
2. Two sleep masks and two sweat bands.
3. Pencils (1 per student).
4. Graphing calculators (1 per student), if regularly used in class.
5. Student worksheets (see page 9).

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

I. Formulate Question(s)
Begin the lesson by providing background information about humans' apparent inability to walk in a straight line without seeing landmarks. Read the following excerpt from the *New Scientist* journal to set the stage:

If we can't see landmarks, we really do end up walking in circles. That's the conclusion of researchers who have tracked people trying to cross pathless deserts and forests. Jan Souman of the Max Planck Institute for Biological Cybernetics in Tubingen, Germany, spends his days researching how the human brain combines information from the senses to build up a picture of the world around it.

When he was approached by a German television station that wanted to track people lost in the Sahara desert for a program about human navigation, he thought the idea was pretty cool. Together they set about investigating the popular belief that people who are lost often find themselves walking in circles.

Souman and his team took volunteers into a forest and the Sahara, fitted them with a GPS tracker and asked them to walk on a straight course through the unfamiliar terrain. When the sun was out, participants managed this surprisingly well, but on cloudy days or at night, with no visual cues to guide them, they veered away, often walking in circles. "I didn't believe that people would really walk in circles, so I was quite surprised," says Souman. (McGinness, 2009)

Then, ask your class if they think that they can walk in a straight line. If you have internet connectivity and a projector, show the associated one minute YouTube video named "We can't help walking in circles" by typing the words "youtube we can't help walking in circles" into your search browser.

In this activity, your students will have an opportunity to test their ability to walk in a straight line while blindfolded. Ask the class if they think that all the students will be able to walk a similar distance prior to going out of bounds. Will there be a large variation in the distances? What might the class distribution look like in terms of distance walked? Discuss this idea and write student predictions on the board.

II. Design and Implement a Plan to Collect the Data
Ask the students what should be considered during the data collection process. For example, noise or talking by peers may give an indication of where the hallway is versus the walls.

For the activity, students will be assigned a partner. One student will walk slowly down the hallway blindfolded with arms slightly extended in front for safety. The other student will accompany the walker and provide an alert when nearing a wall or heading out of bounds. In order to ensure that the walking student cannot see, the sleep mask should be put on first and then the sweatband can be placed on top. Two different pairs of students can walk at the same time with half the class at each end of the 100 feet. Student pairs should write down the distance walked by each member on their student worksheet in the spaces provided (Item 1 on page 9). If possible, each student should have the opportunity to walk down the hallway.

Once back in the classroom, the teacher can compile the data in the table provided on the student worksheet and display it using a document camera. Students should record the complete data set on their own tables (Item 2 on page 9). An example of a classroom set of 20 data points is shown below:

20, 25, 25, 30, 30, 40, 40, 45, 45, 50, 50, 55, 60, 60, 75, 75, 80, 90, 100, 100.

III. Analyze the Data
The first steps in analysis will be to check the normality of the data distribution. Three informal approaches to this process will be used.
   a) The mean will be compared to the median, as the midpoint of the data should be close to the mean.
   b) The median will be compared with the quartiles for symmetry.
   c) A stem plot will be created to see the shape of the distribution and look for outliers.

Assuming the data then is reasonably normal, a t confidence interval for the mean distance walked will be created. If the data is not normal, explain that creating a confidence interval using the student's t distribution is still a valid choice if there are no extreme outliers and the sample size is reasonably large.

a) *Mean and median comparision:*
If your students have calculators, they can load the data into a list. Many calculators and online tools, such as CoreTools, have the ability to provide a 5-point summary under the statistics calculation option. If your students do not have tools available, then have them create a box plot of the data, which corresponds to Item 3 on the student worksheet, page 9. Once students have the box plot, they can find the median and compare it to the computed value of the mean.

b) *Median and inner quartile comparisons:*
Next, the students will find the quartiles in the data. The first quartile (Q_1) is essentially the first quarter of the data. If the number of values between the minimum and median are even, then the average of the two middle values is Q_1, which marks the end of the first quartile and beginning of the second. Similarly, the middle point, Q_3 between the median and the largest value marks the end of the third quartile and beginning of the fourth. Once students have calculated the quartiles, the distance from the median to Q_1 and the distance from the median to Q_3 should be compared. If the distance is much larger on either side, then the data distribution is skewed and not symmetric about the median as we would expect for a normal distribution.

An analysis of the sample data is shown in Figure 1. In this case, the mean and median are fairly close. Comparing the distance between the median and the quartiles, the distance between the median and Q_3 is larger than the distance between the median and Q_1, which indicates that the data is skewed right.

Mean = 54.75 feet   Q1= (40+30)/2   Q3=75 feet
Median = 50 feet   Q1 = 35 feet

Distance from Median to Q1 = 50-35   Distance from Median to Q3 = 75-50
                          15 feet   25 feet

Figure 1. Dotplot and quartiles for the example class data.

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

c) *Stem-plot to view the shape of the data distribution:*
The students should create a stem-plot of the data to see if the sample's distribution is roughly bell-shaped. Each data point should be listed with the stems shown. For example, a data point of 25, would be listed as the numeral "5" with the stem of "2". A completed stem-plot for the same data set as in part b) is shown in Figure 2.
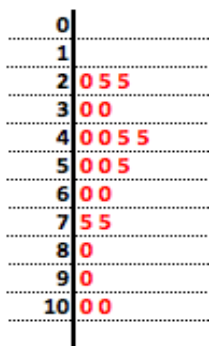
```
 0|
 1|
 2|0 5 5
 3|0 0
 4|0 0 5 5
 5|0 0 5
 6|0 0
 7|5 5
 8|0
 9|0
10|0 0
```

Figure 2. Stem-plot of the example class data.

The benefit of a stem-plot versus viewing a line plot of the data distribution is a consolidation of the data based on the granularity of the selected stems. From the stem-plot, the data does appear to be roughly bell-shaped given the small sample size. With smaller data sets, we would not expect data distributions to look bell-shaped in nature without some grouping of data points.

IV. Interpret the Results
Normality
From the analyses conducted in Item 3, students can now compile all the results regarding the center of the sample data, symmetry, and shape of the sample distribution to determine if the sample is representative of a normal population. Be sure to emphasize that one cannot simply assume that if all the students in the world completed this activity, the distance data would be normally distributed. Clearly, discretion is needed at this point, and a class discussion to determine normality may be the best course of action to arrive at a well-reasoned decision.

Often times, normality is assumed in textbook problems provided to students. Therefore, a discussion of how statisticians determine if a sample is representative of a normal distribution may be a new concept. If normality is not justified through similar comparisons, then statisticians must use different techniques for creating confidence intervals.

For the example data provided, the data appears to be roughly normal given the closeness of the mean to the median and the shape of the stem-plot.

Confidence Interval

As a last step, students can generate the 95% confidence interval for the sample data assuming that normality can be supported. Therefore, students can find the standard deviation for the sample data and the associated approximate confidence interval.

The formula for constructing a 95% confidence interval for the mean distance walked is:

$\bar{x} \pm t_{.025} \left(\dfrac{s}{\sqrt{n}} \right)$, where $\bar{x}$ is the sample mean distance walked, s is the sample standard deviation,

and n is the size of the data sample. For a 95% confidence interval, the $t_{.025}$ value can be found in a table online, such as: http://www.medcalc.org/manual/t-distribution.php or via an online applet, such as: http://www4.stat.ncsu.edu/~west/applets/tdemo.html (Java needs to be enabled) or from the invT function on a TI calculator. The degrees of freedom are determined by $df = n-1$, and we are interested in the t-value associated with the 2.5% regions on the right and left tails of the t distribution. If your students don't have access to a table, Internet applet, or the invT function, you may want to provide this value to them and explain what it represents.

For the sample data provided, the associated 95% confidence interval computation becomes

$54.75 \pm 2.093 \left(\dfrac{24.84}{\sqrt{20}} \right)$. The resulting confidence interval has a lower endpoint of 43.12 feet and an upper endpoint of 66.38 feet. Thus, we are 95% confident that the mean distance students can walk before going out of bounds is between 43.12 feet and 66.38 feet.

Be sure to discuss what this interval means, i.e. 95% of all confidence intervals constructed in the same manner would include the population mean distance walked for all similar students. 5% of them would miss or not include the mean. Extensions to other populations, such as all people, is inappropriate, as the sample data were taken from a specific group.

Assessment

Quinn and her classmates obtained the following data when walking blindfolded down a 100-yard hallway in her school.

```
 0|5
 1|0055
 2|0055
 3|05555
 4|000055
 5|0005
 6|55
 7|00
 8|05
 9|
10|
```
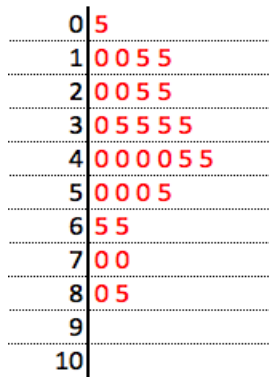
Quinn says, "Obviously, the sample data indicates that the underlying population is normally distributed. You can tell from the bell-shape of the stem-plot".

1) What other two attributes of the sample data does Quinn need to check before she assumes the population data is normally distributed?

    a) _____   b) _____

2) Check these aspects and determine if normality can be assumed. Be sure to state why or why not in your response.

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

Answers

1) a) Compare the mean and median values.
   b) Compare the distance between the quartiles and the median.

2) mean = 40.2, median = 40. These are very close and support the normality assumption.

   Q1 = 25, median – Q1 = 15, Q3 = 50, Q3 – median = 10. Since these are close to the same, the distribution appears to be symmetric.

Possible Extensions

Depending upon class size, the teacher could separate the data for boys and girls and perform a comparative analysis on the walking distances. Descriptive statistics could be calculated for each group. Further, a significance test could be performed to compare the mean walking distances of the two groups. Additionally, a confidence interval could be constructed in order to estimate the difference between the mean walking distance for boys and girls.

References

1. Adapted from an activity created by Daniel A. Kaufmann, Wilson High School, West Lawn, Pennsylvania. http://goo.gl/zODZ3

2. McGinness, Laura (2009). We can't help walking in circles. *New Scientist.* http://www.newscientist.com/article/dn17658

Walk the Line Activity Sheet

Name:…………………………………………   Date: …………………

1. Write the distance that you and your teammate travelled down the hallway before going out of bounds.

_____ feet   _____ feet

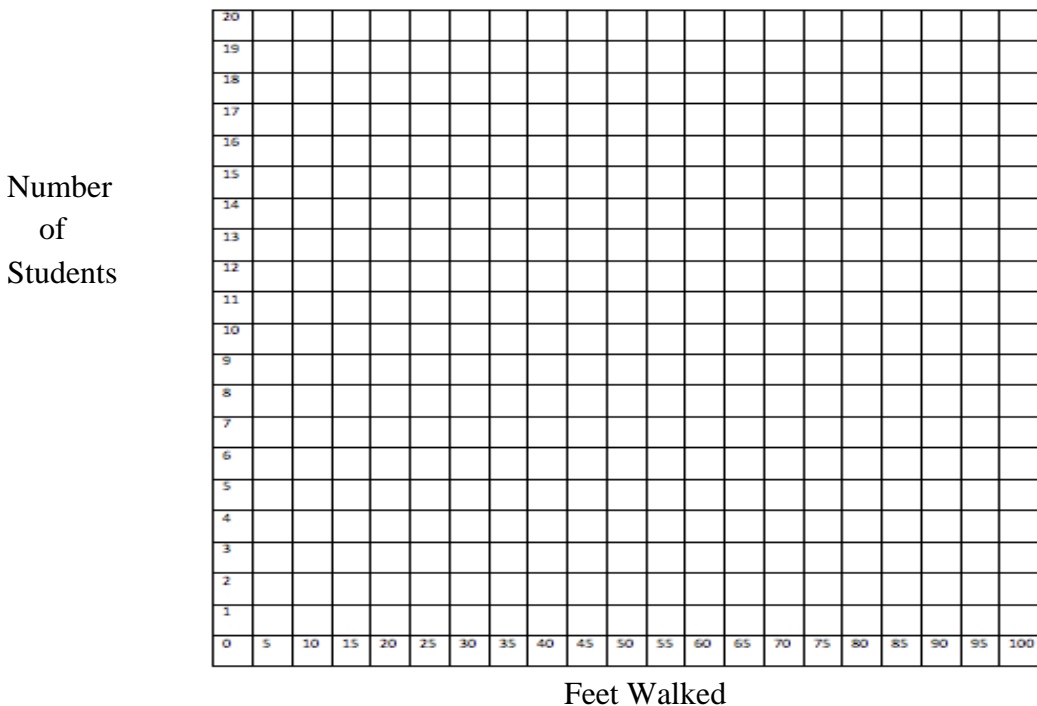**2.** Use this table for the full class data set.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
21.	22.	23.	24.	25.	26.	27.	28.	29.	30.
31.	32.	33.	34.	35.	36.	37.	38.	39.	40.

Normality Check

3. Generate a line plot of the class data, find the mean, median and quartiles.

Data Distribution of Distance Walked

Number
of
Students

Feet Walked

Comparison of Mean and Median

a) What is the mean value? _____   Median value?_____

Are these two values approximately the same and explain why?   Yes or No (circle one).

Explain why_____

Symmetry Check

b) Determine if the data are symmetric about the mean

i) What is the "feet walked" value that marks the beginning of the second quartile?_____

ii) What is the distance from the beginning of the second quartile to the median?_____

iii) What is the "feet walked" value that marks the beginning of the fourth quartile?_____

iv) What is the distance from the beginning of the fourth quartile to the median?_____


v) Are these two distances approximately the same in parts ii and iv?   Yes or No (circle one).

Explain why_____

If the distances are approximately the same, then we can say that the data distribution is roughly symmetric.

Shape of the Data Distribution

c) Generate a stem-plot to assess the overall shape of the data distribution. Using the data for your entire class, create a stem plot below to show the shape of the distribution.

```
 0|
 1|
 2|
 3|
 4|
 5|
 6|
 7|
 8|
 9|
10|
```
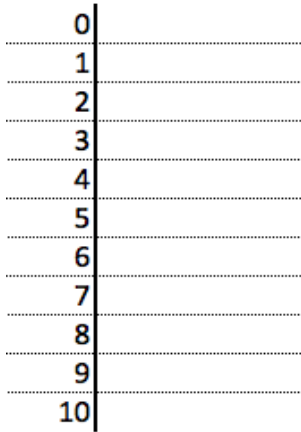
Is the data distribution approximately bell-shaped?   Yes or No (circle one).

Explain why or why not_____

4. In testing sample data for normality or "likeness to the normal distribution", we can compare the mean to the median, seek symmetry, and view the overall shape of the sample distribution.

From the results of your analyses, do the results support normality of the sample data?
Yes or No (circle one).

Explain why or why not_____

5. Assuming that the data set for the class is approximately normal, construct an approximate 95% confidence interval for the mean distance students can walk down the hallway before going out of bounds.

   a) Find the standard deviation for the data set: $s =$ _____

b) The formula to construct a 95% confidence interval for a population mean using a student's *t* distribution is:

$\bar{x} \pm t_{.025} \left(\dfrac{s}{\sqrt{n}} \right)$, where $\bar{x}$ is the sample mean, *s* is the sample standard deviation, and *n* is the size of the data sample. For a 95% confidence interval, the $t_{.025}$ value corresponds to the probability that the value is larger than 97.5% of the other values.

95% confidence interval = (_____,_____)

c) What does this interval represent?_____

_____

d) Will the interval created be representative of *all* students?   Yes or No (circle one).

Explain why or why not_____

e) To what population will the results match most closely?_____