Normal Distributions

Is Fall Normal?

I. ASSESSSMENT TASK OVERVIEW & PURPOSE:

Students will collect 100+ leaves and a large sample of a second object of their choosing. Students will measure the length of the leaves from the base of where the stem meets the leaf to the longest point of the leaf in centimeters. The second object must be something that occurs in nature and preferably something specific to the fall season. Examples of such items could include specific vegetables such as squash, pumpkin, pumpkin seeds, pine cones, pine needles, and acorns. They will examine a measureable feature of the object such as its mass or length. Students will organize their data in a spreadsheet software such as MS Excel. Students will calculate the descriptive statistics and determine if the data sets are normally distributed. Students will be encouraged to use Empirical Rule as a means of assessing the normality of the data sets in addition to other features of normal distributions.

II. UNIT AUTHOR:

Melissa Harnist
Roanoke Valley Governor's School
Roanoke City Public Schools

III. COURSE:

Algebra II

IV. CONTENT STRAND:

Algebra II, Statistics

V. OBJECTIVES:

The learner will be able to:

- Define and apply the terms relating to quantitative and qualitative data.
- Calculate and interpret the appropriate measures of central tendency for a set of data.
- Calculate and interpret the appropriate measures of variation for a set of data.
- Construct and interpret box-and-whisker plots.
- Identify the properties and characteristics associated with any normal distribution.
- Label a graph showing the critical data for a standard normal distribution.

VI. REFERENCE/RESOURCE MATERIALS:

- 100 leaves collected by the class from an obliging tree
- 50 natural "fall" related items brought in by each pair of students
- 3-4 balances (1 balance for every 5 students would be a reasonable ratio)
- 10 - 20 rulers (1 per student or enough for 1 per pairing)
- Classroom set of computers with MS Word (at least 1 computer for each pairing of students)
- MS Excel or Other Statistical Software such as Fathom or Minitab

VII. PRIMARY ASSESSMENT STRATEGIES:

Students will create a lab write-up in an MS Word document that incorporates the computations and graphics created in MS Excel. Students will articulate their interpretations of the data set and their findings regarding the normality of the data set. A handout and a rubric which outlines the components of their write-up is included with this lesson template.

Students will complete this assignment with a partner. This will help with any accommodations that are needed for students with special needs.

VIII. EVALUATION CRITERIA:

Students will produce a type-written report that includes:

- An introduction that describes the purpose of the activity and the procedures of the lab, with an emphasis on the strategies used to randomly collect the data
- Two separate data tables containing the descriptive statistics for the leaves and the second object of their choosing
- A description of the average leaf length and the average mass/length of their second object.
- A description of the amount of the amount of variance within each data set.
- An assessment of whether or not the two data sets follow Empirical Rule.
- An explanation of the symmetry of the data set that is supported with some graphical display.
- A conclusion regarding whether or not the length of leaves is normally distributed and whether or not the second data set is normally distributed.
- Effective and correct, writing, grammar, and submission techniques.

IX. INSTRUCTIONAL TIME:

Two 90 minute periods.

Is Fall Normal?

Strand
Algebra II, Statistics

Mathematical Objective(s)

Students will demonstrate mastery of their knowledge of normal distributions as they assess if a set of data that they collect fits the criteria of a normal distribution. More specifically, students will:

- Define and apply the terms relating to quantitative and qualitative data.
- Calculate and interpret the appropriate measures of central tendency for a set of data.
- Calculate and interpret the appropriate measures of variation for a set of data.
- Construct and interpret box-and-whisker plots.
- Identify the properties and characteristics associated with any normal distribution.
- Label a graph showing the critical data for a standard normal distribution.

Furthermore, students will demonstrate proficiency with the use of statistical software to facilitate the organization of and analysis of data as they use MS Excel and MS Word for creating the graphics and text for their report.

Related SOL
AII.11 (Normal Distributions)

NCTM Standards

- Understand the meaning of measurement data and categorical data, of univariate and bivariate data, and of the term variable;
  Understand histograms, parallel box plots, and scatterplots and use them to display data;
- Compute basic statistics and understand the distinction between a statistic and a parameter.
- For univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics;
- Apply and adapt a variety of appropriate strategies to solve problems
- Communicate mathematical thinking coherently and clearly to peers, teachers, and others

Additional Objectives for Student Learning:
BIO 1.d,e (Data analysis, quantitative data)

Materials/Resources

- 100 leaves collected by the class from an obliging tree
- Large sample of second object of the students' choosing (50+ items preferred)
- 3-4 balances (1 balance for every 5 students would be a reasonable ratio)
- 10 - 20 rulers (1 per student or enough for 1 per pairing)
- Classroom set of computers with MS Word (at least 1 computer for each pairing of students)
- MS Excel or Other Statistical Software such as Fathom or Minitab

*Be sure that all leaves are collected from the same kind of tree. Consider having students bring their second set of objects to school to measure their length or mass as appropriate. These items may be collected and stored in the classroom ahead of time if it fits the timeline of instruction better. Balances and rulers will need to be available from the start of the activity.

Assumption of Prior Knowledge

- *Prior mathematical knowledge assumed by the task/activity.*
  Students should have a functional level of knowledge regarding methods for calculating and interpreting descriptive statistics. Students need to be familiar with defining characteristics of normal distributions such as how the mean, median and mode relate, the symmetry that occurs with normal distributions, and Empirical Rule.

- *Student background knowledge*
  The activity that students will be doing in this lesson involves students performing procedures with connections and also "doing" the mathematics. Therefore, students should be proficient with identifying properties of normal distributions, working with standard deviations, and creating and using graphical displays of data to investigate symmetry.

- *Common language, ideas, and terminologies used*
  Students should be using terminology such as normal distributions, standard deviation, mean, median, mode, range, interquartile range, minimum, maximum, symmetry, skewness, and outliers.

  Students should pay special attention to the how the mean, median and mode of each data set relate to one another. They should be comfortable with ways to assess the symmetry of the data such as a boxplot, a frequency or dot plot, or a stem-and-leaf plot. Students should be familiar with Empirical Rule and be able to determine the number of data values that fall within one, two, and three standard deviations of the mean.

- *Possible misconceptions or challenges*
  Students sometimes become too tedious in this activity that they lose sight of the bigger picture in determining if the leaves and pinecones which they collected are normal. In general, most data sets collected from nature are normally distributed. However, some students become overly concerned that the leaves may have a mean length of 7.5cm and a median length of 7.4cm and they decide that the data set is not normal when this should not be enough of a difference to deter it from being normal.

- *Relevant concepts have already been (or should have been) explored prior to this task/activity*
  - Methods for calculating and interpreting descriptive statistics
  - Methods for creating graphical displays of data that show the spread of the data (left skewed, right skewed, or symmetrical)
  - Characteristics of normal distributions
  - How to use MS Excel or other statistical software to find descriptive statistics

- *What relevant contexts (example: analysis & impact of natural disasters; traffic control; social issues) are drawn on in relation to this concept?*
  Students are able to apply their knowledge of quantitative and qualitative data that are often discussed more heavily in the science classrooms than in the math classrooms. Furthermore, students are able to make connections to what types of data sets can they expect to be normal.

Introduction: Setting Up the Mathematical Task

In this lesson, students will collect and analyze data regarding the length of 100 leaves and collect a second object of their choosing and analyze its length or mass. Students should obtain prior approval of their second object prior to the start of this lesson. Examples of such objects could include but are not limited to: mass of pumpkin seeds, length of pumpkin seeds, length of pine needs, mass or an acorn, mass of a pinecone, length of a squash, or the mass of squash. Students will need to have access to a large sample size of this second object. It is recommended that students have access to at least 50 individual data points for this second item. Students will apply their knowledge of normal distributions to assess if the data which they collected is normal.

Time outline: 2-90 minute periods:

- 10 minutes - The teacher will facilitate a classroom discussion to develop methods for data collection. This should include specifying how many leaves will need to be collected by each group. Students should look for "whole" leaves with few "holes" and pinecones that are complete and not missing any of its chippings.
- 55 minutes - Students will need to go outside and collect leaves. Students will need to have brought in their second item for class this day. They will take the appropriate measurements for each object: the length of the leaves and the mass or length of their second object. Students will want to record the leaf lengths in a way that allows all groups to share the data. The teacher can facilitate this process through the use of a Google document or having students enter their data on one central computer that the teacher then sends out the class data to the students to analyze further in their groups.
- 10 minutes - The teacher will need to help redistribute all of the class' data for the leaves to the students' groups. All groups need to have access to all 100 leaves' lengths. Students will analyze their group's second object.
- 55 minutes - Students will need to use spreadsheet software such as MS Excel to sort, organize, and create appropriate graphical figures to help them interpret their data sets.
- 45 minutes - Students will compile their findings in a type-written report.
- 5 minutes - The teacher will synthesize each group's conclusions to ensure that all groups arrived at the appropriate conclusions regarding the leaves and their second data sets.

Student Exploration

Whole Class Sharing/Discussion
The teacher will facilitate a class discussion prior to data collection to ensure that students implement the same strategies and criteria for gathering the leaves and their second object. The leaves need to be measured in the same way and the students' second set of objects also need to be measured using consistent methods.

Small Group Work
Students will gather the leaves and their second set of objects in groups and take the appropriate measurements in groups. Students may have groups of up to four students for this task; however, they will organize and analyze their data with a partner. Students should touch base with their teacher after creating the appropriate graphs and performing the necessary calculations for this lab. The teacher will have an informal discussion with students to ensure that they draw reasonable conclusions from their graphs and tables.

Student/Teacher Actions:

- After gathering all the data, students should sort each data set separately, in ascending order. They will need to find the descriptive statistics (mean, median, mode, range, standard deviation, quartile one, quartile two, quartile three, the minimum value, and the maximum value). The students should count the number of data values that fall within one, two, and three standard deviations of the mean data value and convert these numbers to percentages. Students should compare these percentages with the expected percentages that correspond with Empirical Rule. Students should create a graphical display to examine if each data set is symmetrical. These graphical displays could be a box plot, a dot plot or frequency plot, or a stem-and-leaf plot. Students should check with their teacher for an informal discussion regarding if each data set is normal after they have created a graphical display for each data set and a table organizing the percentage of the data values that fell within one, two, and three standard deviations of the mean for each data set.

- The teacher should ask guiding questions to help students through each of these tasks and help troubleshoot possible technology concerns. Some students may be comfortable using other statistical software if it is available. Most statistical software packages will be capable of handling the demands of this activity.

- Teachers should be watchful for the following common misconceptions: students tend to expect to see "exact" matches without seeing general patterns. Some students become too focused on wanting the key features of normal distributions to line up perfectly. If the mean, median and mode are approximately the same and the percentage of data values that fall within one, two and three standard deviations are close to those that should observed with normal distributions then it is enough to conclude that the data set is normal. The reality is that 50 - 100 data values is a large sample for the students to work with, but it is not a large sample when thinking in the broader context of statistics. To help students see this further, if multiple classes conduct this activity, consider merging the data sets of all the classes for the leaves and show the students the results at the conclusion of this lesson.

Monitoring Student Responses

Student responses will be monitored through four distinctive methods for this exploration:

1. Students' discussions with their partner or groups as they gather and analyze their data for the leaves and pinecones.
2. Students' reports where they justify their reasoning for if each data set is normal
3. Classroom discussion that is facilitated by the teacher prior to the data collection process and before students analyze all the combined class data.
4. Teacher discussion with each pairing of students after they create the graphs and tables required for their write-up.

Assessment List and Benchmarks

Student handout containing general rubric:

Is Fall Normal?

Due:
100 Points

Statistics can be used to describe and analyze naturally occurring phenomenon. This knowledge can help scientists recognize significant changes in a given population. Your goal in this activity is to apply your knowledge of statistics to determine whether or not the length of leaves and the mass or length of a second "fall item" of your choosing are normally distributed data sets.

Materials:

- Balance
- Ruler
- 50 total items for a second data set
- 100 total leaves
- Computer with spreadsheet or statistical software

Procedures:

1. In the area designated by your instructor, collect whole leaves. Make sure that you collect only one kind of leaf (i.e. from the same kind of tree) and that you collect only whole (or mostly whole) pine cones.
2. Determine a second "fall" item of your choosing that occurs naturally and obtain a sample size of at least 50 items of this object. This task may be divided between you and your partner. Bring these items to school on the day designated by your instructor. Examples of "fall" items might include - pumpkin seeds, pine needles, or acorns.
3. Divide the 100 leaves among the groups so that each group is responsible for relatively the same amount of leaves for data collection. Measure in centimeters the length of each leaf from the base of where the stem meets the leaf to the tip of the leaf and record the length.
4. As a class, determine the best way to compile the data so that all groups will have access to the data pertaining to all 100 leaves.
5. Have each group measure the mass or length of their second "fall" data set and enter these values into a spreadsheet.

Statistical Analyses:

1. Find the descriptive statistics for the leaves and the second object. Organize these values in a data table.
2. Examine each data set for symmetry or skewness (right or left skew) and have a graphical display to support your conclusion.
3. Determine whether or not each data set follows Empirical Rule. Create a table that organizes the actual percentages that fall within the appropriate standard deviations from the mean. Use this table to help with your written justification of your conclusion.

The Product:

A typed report with a partner that includes:

- An introduction that describes the purpose of the activity and the procedures of the lab, with an emphasis on the strategies used to randomly collect the data
- Two separate data tables containing the descriptive statistics for the leaves and the second object - Use the values in the data table to describe the average data value and the amount of variance within each data set
- A graphical display to illustrate the symmetry or skewness for each data set - Explain the information shown in the graphical display and explain how this relates to the normality of the data.

- Two data tables used to compare each data set to the percentages that correlate with Empirical Rule.
- Conclusions regarding whether or not the length of leaves is normally distributed and whether or not the second set of items is normally distributed. ***Justify your answer based upon the defining features of normal distributions.***
- Effective and correct, writing, grammar, and submission techniques.

Is Fall Normal? Rubric

Criteria	Points Earned	Points Possible
Introduction • Clearly describes the purpose of the activity and the procedures of the lab - 3 points • Includes a brief discussion of the strategies used to randomly collect the data - 2 points		**5**
Data tables of descriptive statistics • Includes measures of central tendency and measures of variance, - mean, median, mode, standard deviation, and the five number summaries for each data set - 12 points • Data tables are organized in a way that is meaningful and clear - 3 points • Values are calculated correctly. - 5 points		**20**
Description of the "average" value and the amount of variance in each data set. • Selects and uses an appropriate descriptive statistic to describe the "average" data value - 5 points • Uses measures of variance to describe how the data falls in relation to the average value. - 5 points		**10**
Graphical displays to illustrate the symmetry or skewness of each data set [A total of two, one for each data set] • An appropriate graphical display to show the spread of the data is selected for each data set - 6 points • Graphs contain a title - 2 points • Axes are labeled appropriately - 2 points		**10**
Explanation of the each graphical display • Interpretation of what each display shows • Appropriate connections drawn between the display and whether or not it supports the idea that the data set is normally distributed.		**20**

Empirical Rule Comparisons • A total of two data tables are present containing the actual percentages for the appropriate standard deviations - 5 points • Connections are made between these data tables and Empirical Rule - 5 points		**10**
Conclusions regarding normality • Conclusion regarding whether or not the length of leaves is normally distributed is justified using multiple features of normal distributions - 10 points • Conclusion regarding whether or not the second "fall" item is normally distributed is justified using multiple features of normal distributions - 10 points		**20**
Writing and Grammar • Few or no typos are contained in the document. - 2 points • Proper grammar is used. - 1 point • Proper mathematical vocabulary is used. - 2 points		**5**
Total		**100**

Actual sample student work:
*Note, her class accidentally miscounted the leaves and used 99 leaves. Her group used 83 pinecones. Even though these values differ from the suggested size of 100 and 50, as long as the size of the data sets are both sufficiently large, students should be able to make the appropriate connections.

Is Fall Normal?

Cameron *****

The purpose of this activity was to prove how everyday people could use statistics to describe naturally occurring phenomenon. The event that was analyzed in this particular activity was the changing of trees during fall. In order to determine if fall is normal, 99 leaves and 83 pinecones were collected randomly from outside of Roanoke Valley Governor's School. The lengths of the leaves and mass of the pinecones were measured and then analyzed to see if it was normally distributed using descriptive statistics, box plots, and variance. It was very important to ensure that the leaves were all from the same type of tree because different trees produce leaves of a variety of shapes and sizes, thus changing the length. However, it was also important that the leaves were not all collected from the same tree, because the condition of the tree is unknown. It is very possible that the tree could be dying, which would affect the length of its leaves. These same concepts were taken into account while collecting the pinecones. All of the pinecones used for analysis were whole so that the mass would not be affected.

Once all of the leaves and pinecones were measured for length (leaves) and mass (pinecones), the descriptive statistics of each of the data sets were determined. Below is the descriptive statistics of the leaves.

Leaf Length (cm)	Mean	Median	Mode	Range	Standard Deviation	Quartile 1	Quartile 2	Quartile 3	Max	Min
	10.1071	10.1	10	6	1.52872	9	10.1	11.1	13	7

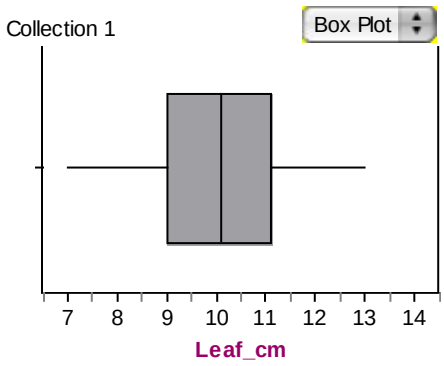
From subtracting Quartile 1 from Quartile 3, the Interquartile Range (IQR) was determined to be 2.1. The data is spread between 7 and 13 cm. The data is very consistent, although it is not perfect. The average of the length of the leaves is best shown by the median, since exactly 50% of the leaves collected are shorter than 10.1 cm and 50% are longer. The same cannot be said about the mean. The descriptive analysis of the pinecones is shown below.

Pinecone Mass (g)	Mean	Median	Mode	Range	Standard Deviation	Quartile 1	Quartile 2	Quartile 3	Max	Min

	4.39036	4.25	N/A	5.313	1.35929	3.203	4.25	5.42	7.31	1.997

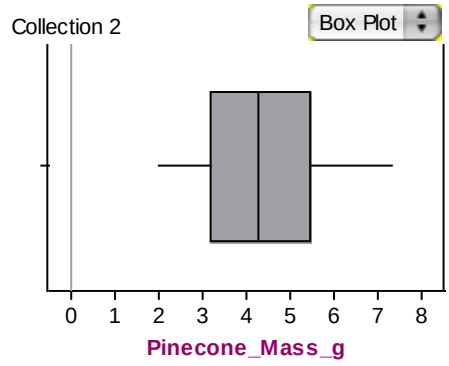
The IQR of this data was found to be 2.217 using the same method of subtracting Quartile 1 from Quartile 3. It should be noted that no two masses of the pinecone were equal; therefore there was no mode. The pinecone masses were spread from 1.997 to 7.31 g. The average of this data should not be represented by the mean, because it is not true that exactly 50% of the data weighs less than 4.39036 g. and 50% of the data weighs more. Although the mean and median are very close, the median most accurately describes the average of this data.

The box plot of the length of the leaves is shown here. The five-number summary associated with this graphic



was listed in the descriptive statistics above, but again the minimum value was 7, quartile 1 was 9, quartile 2 was 10.1, quartile 3 was 11.1, and the maximum value was 13. This means that 50% of the data is more than 10.1 and 50% of the data is less than 10.1. Also, 25% of the data is between 7 and 9, 9 and 10.1, 10.1 and 11.1, as well as between 11.1 and 13. Finding the median of the minimum value and quartile 2 shows the value of quartile 1. Similarly, quartile 3 is the median of quartile 2 and the maximum value. The excellent thing about using a box plot and five-number summary in statistics is that the skew can be determined relatively easily. The mean and median of this data are not the same. The median (also known as quartile 2) was found slightly to the right of the middle of the shaded box in the box plot. That means that the data is ever so slightly left-skewed.

Now, this is the box plot for the mass of the pinecones. The five-number summary was also stated in the



descriptive statistics, but the minimum value was 1.997, quartile 1 was 3.203, quartile 2 was 4.25, quartile 3 was 5.42, and the maximum value was 7.31. Half (or 50%) of the pinecones weigh less than the median (4.25) and 50% weight more. Also, 25% of the data is between 1.997 and 3.203, 3.203 and 4.25, 4.25 and 5.42, as well as between 5.42 and 7.31. The median of the data was slightly on the left of the middle of the shaded box on the box plot, which means that the data is barely right skewed. The reason for the skew is that there is more data concentrated around a mass of 4.25 g.

Shown below are the percentages of data that falls within 1, 2, and 3 standard deviations of the mean for the length of the leaves.

Within one standard deviation of the mean	63.64%[1]
Within two standard deviations of the mean	98.99%[1]
Within three standard deviations of the mean	100%

This was found by first aligning all of the data in ascending order using Microsoft Excel. The standard deviation found in the descriptive statistics was subtracted from the mean, and then the standard deviation was added to the mean. These values were found to be 8.57838 and 11.63582, respectively. This same method was used to find the values two and three standard deviations below and above the mean. Next, the values just calculated were used to find the percentage of the original data that falls between these values. Since the data was organized in Excel, the cell numbers in which the data lied were subtracted from each other and then divided by the total number of data points, in this case, 99. This value was turned into a percentage by being multiplied by 100. The same process was used to find the percentage values of the data that lies within one, two and three standard deviations of the mean mass of the pinecones. These values are listed below. Please note that these percentages are also rounded to the nearest 100^{th} place.

Within one standard deviation of the mean	62.65%
Within two standard deviations of the mean	97.59%
Within three standard deviations of the mean	100%

[1] The percentage values are rounded to the nearest 100^{th} place

Based off of the information in this lab, it can be concluded that the leaves are normally distributed. One of the main characteristics of normal distributions is that the mean, median, and mode are the same. It is understood that there was some error that occurred in the lab; therefore the results are not perfect. However, the mean, median and mode of the length of the leaves were all very close to 10 cm. As for the mass of the pinecones, the difference between the mean of 4.39 g and the median of 4.25 g is not noticeable to the touch. While it may seem that the fact that there was no mode threw off the chances of it being normally distributed, it must be noted that there were approximately 5 values that

were around 4.3 g. Another characteristic of normal distribution is that there is no skew. While it was said earlier that there was a skew, it was also stated that the skew was very slight. That skew could have easily been caused by a small error made while recording the measurements. One last defining feature of a normal distribution is that it follows Empirical Rule. This means that approximately 68% of the data lies within one standard deviation, 95% within two standard deviations, and 99.7% within three standard deviations. While Empirical Rule is an excellent tool, the values are also approximations. The percentage values found in the tables above are very close to following Empirical Rule, especially considering that the 68%, 95%, and 99.7% values are approximations and there was almost surely human error. Also, the data used in this lab were samples for all leaves and pinecones. For all of the reasons listed above, it can be concluded that fall is normal.