Math Objectives

- Students will recognize that the mean of a skewed data set may not be a useful measure of center.
- Students will compare different transformations of univariate data and determine which one is more effective at eliminating skew.
- Students will understand that univariate data can be transformed in order to apply z-scores and inference procedures.
- Students will construct viable arguments (CCSS Mathematical Practices).
- Students will look for and make use of structure (CCSS Mathematical Practices).

Vocabulary

- exponentiation
- logarithm
- mean
- skew (left and right)
- square root
- symmetric
- univariate data

About the Lesson

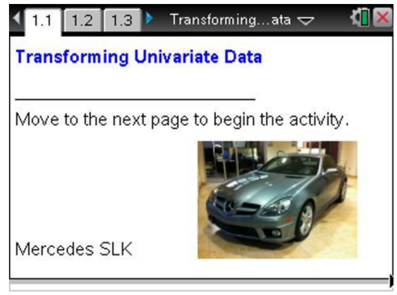
- This lesson involves square root, logarithmic, square, and exponentiation transformations of skewed univariate data using a given data set.
- As a result, students will:
  - Look at data that has been transformed using square roots and logarithms and determine which is more effective at reducing skewness in a given distribution.
  - Look at data that has been transformed using squares, exponentiation, square roots, and logarithms and determine which is more effective at reducing skewness in a given distribution.
  - Recognize that performing data transformation can eliminate skewness in a distribution, allowing use of the transformed mean and standard deviation in z-scores.

TI-Nspire™ Navigator™ System

- Send out the *Transforming_Univariate_Data.tns* file.
- Monitor student progress using Class Capture.
- Use Live Presenter to spotlight student answers.

Activity Materials

- Compatible TI Technologies: TI-Nspire™ CX Handhelds, TI-Nspire™ Apps for iPad®, TI-Nspire™ Software



Tech Tips:

- This activity includes class captures taken from the TI-Nspire CX handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

Lesson Files:

Student Activity

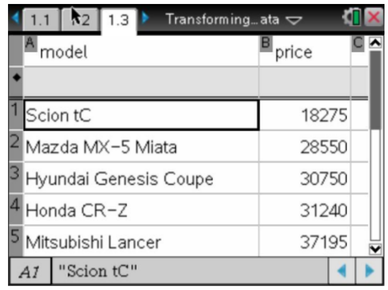
- Transforming_Univariate_Data_Student.pdf
- Transforming_Univariate_Data_Student.doc

TI-Nspire document

- Transforming_Univariate_Data.tns

Discussion Points and Possible Answers

Move to page 1.3.

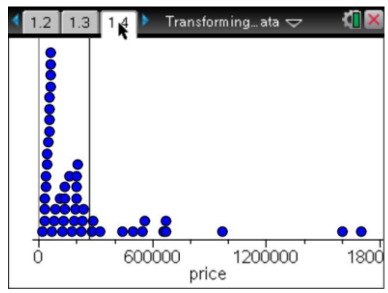


The data on Page 1.3 are list prices for 2011 sports cars.

1. Scroll through the list of prices. What observations can you make?

   <u>Sample Answer:</u> Two of the cars cost over $1.5 million dollars, really expensive.

Move to page 1.4.



Page 1.4 shows a dot plot of the car data with a vertical line at the mean. If you select the line, the value of the mean is displayed.

2. a. Describe the distribution of sports car prices with respect to shape and center. Does the distribution surprise you given what you saw in the spreadsheet?

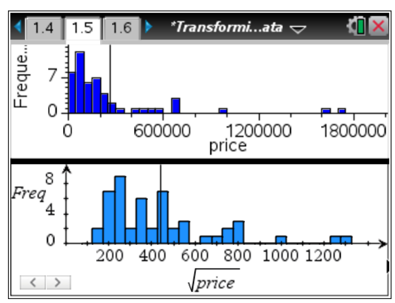
   <u>Sample Answer:</u> The distribution of sports car prices is strongly skewed to the right with a mean of $263,315.

b. Where does the mean fall within the list of data values?

   <u>Sample Answer:</u> The mean is larger than all but 13 of the 50 values.

> **Tech Tip:** Students can grab and drag the x-axis to the left and right to see additional intervals.

Move to page 1.5.



This page contains histograms of the skewed distribution of the sports car prices. We want a transformation that changes numbers in ways that makes the distribution more symmetric and mound-shaped; therefore, we will explore both square root and logarithmic transformations.

3. Press the arrow in the lower work area to take the square root of all of the car prices. The vertical lines show the means of the distributions.
   a. How does the distribution change with respect to shape? Where does the mean of the transformed data fall within the list of data values?

      Sample Answer: The shape of the distribution of car prices is less skewed than the distribution of the original data but still not very symmetric; the mean of the square root transformed car prices is 439.39, which is larger than all but about 20 of the 50 values

   b. What are the units of the transformed data?

      Sample Answer: Squareroot(dollars).

   c. If you were to use the transformed data to get a more representative indicator of center for the original data, what would you need to do to the transformed mean in order to have the original units?

      Sample Answer: You would need to square the value to “un-transform” the mean and get a value in dollars.

4. Another way to change the shape of the distribution is to take their common (base 10) or natural (base e) logarithm. Press the right arrow again to take the common logarithm of all the car prices. The vertical lines show the means of the distributions.
   a. How does the distribution change with respect to shape? Where does the mean of the transformed data fall within the list of data values?

      Sample Answer: The shape of the distribution of car prices is now approximately symmetric; the mean of the log-transformed data is 5.15142, which appears to be in the middle of the distribution.

   b. What are the units of the transformed data?

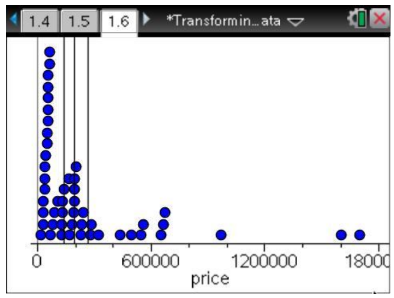
      Sample Answer: Log(dollars).

   c. If you were to use the transformed data to get a more representative indicator of center for the original data, what would you need to do to the transformed mean in order to have the original units?

Sample Answer: You would need to raise ten to the mean of log(dollars) power to "un-transform" the mean and get a value in dollars.

d. Which transformation made the data more symmetric and mound-shaped? Explain your reasoning.

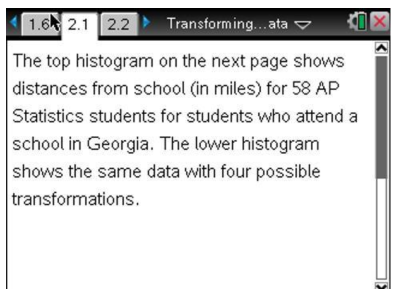
Sample Answer: Taking the log of the prices produced the most symmetric distribution. The mean is larger than 26 of the values and smaller than 24, close to the middle of the transformed prices.



Move to page 1.6.

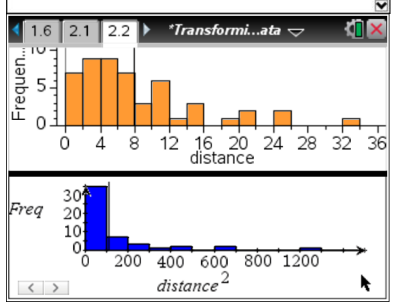
5. In order to put the mean back into the original units, in 4c you had to square it, and in 5c you had to exponentiate it. The three vertical lines in on Page 1.7 show the mean of the original data and the back-transformed means of the other two transformations. Which measure of center seems most representative of all of the data? Explain your reasoning.

Sample Answer: The log transformed mean ($141, 717.) seems most representative—it is larger than 26 of the values and smaller than 24, close to the center of the prices.



Move to page 2.1.

This page introduces another data set. These data are the distances in miles from school for 58 AP Statistics students who attend a school in Georgia.



Move to page 2.2.

6. The top histogram displays the distance data with a vertical line at the mean. If you click on the line, the value of the mean is displayed. Describe the distribution of distances with respect to shape and center.

Sample Answer: The distribution of distance is skewed to the right with a mean of 7.82 miles.

7. Use the arrow in the lower work area to check four different transformations of the distance values (square, exponential, square root, and logarithmic). Which transformation made the data the most symmetric and mound-shaped? Explain your reasoning.

   Sample Answer: Taking the square root of the distances produced the most symmetric, mound-shaped distribution. Squaring and exponentiating the data values made the distributions even more skewed to the right. The logarithmic transformation "over-corrected" and made the distribution slightly skewed to the left.

Wrap Up

Upon completion of this lesson, teachers should ensure that students are able to understand:

- That the mean of a skewed univariate data set may not be a useful measure of center.
- That skewed univariate data can be transformed to make the distribution more symmetric and mound-shaped by squaring, raising to a power, 'squarerooting,' or taking the common (or natural) logarithm.
- The need to "un-transform" transformed data.

Assessment

Label each of the following as sometimes, always, or never. Be prepared to defend your answers.

1 – In a strongly skewed distribution, the mean will be pulled in the direction of the skewness.
Sample Answer: always

2 – Square root transformations will reduce skewness in a distribution.
Sample Answer: sometimes

3 – Logarithmic transformations will reduce skewness in a distribution.
Sample Answer: sometimes

4 – The mean is a good measure of center in a distribution of data.
Sample Answer: sometimes