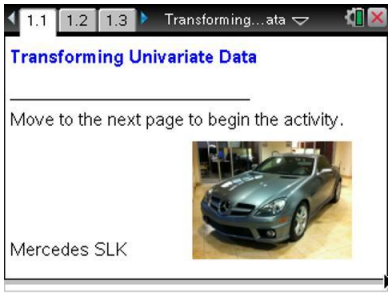
Transforming Univariate Data

Student Activity

Name _____

Class _____

Open the TI-Nspire document *Transforming_Univariate_Data.tns*.



Mean, standard deviation, z-scores, and all of the inference techniques based on z-scores work best when data distributions are reasonably symmetric and mound-shaped. Real data frequently do not cooperate. Our goal is to transform (or re-express) univariate data values to address the problem of skewness.

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?

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?

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

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?

b. What are the units of the transformed data?

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?

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?

   b. What are the units of the transformed data?

   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?

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

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.

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.

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.

This page intentionally left blank.