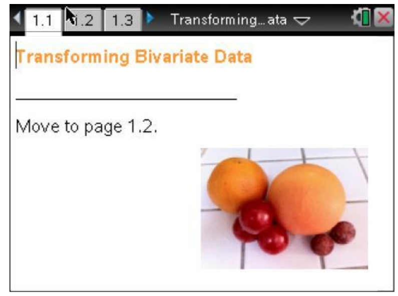
Transforming Bivariate Data
Student Activity

Name ______________

Class ______________

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



Linear relationships are important in analyzing data. In this activity, you will explore strategies for transforming data to see if it is possible to find a linear relationship between the transformed values.

Analyzing the relationship between two variables, usually an explanatory variable and a response variable, is important in statistics. Statisticians have developed tools to help them accomplish this task. The theory and assumptions necessary to use these tools, however, are based on the use of linear relationships between the variables. Unfortunately, many relationships are not linear. This activity explores techniques for transforming non-linear relationships into linear relationships.

Move to page 1.3.

Press ctrl ▶ and ctrl ◀ to navigate through the lesson.

The data on Page 1.3 are the type, the circumference in centimeters, and the mass in grams of 14 approximately spherical pieces of fruit.

1. Use the bar on the right to scroll through the measurements. What do you think a scatterplot of these bivariate data will look like?

Move to page 1.4.

Page 1.4 displays a scatterplot of the fruit data with the least-squares regression line and the correlation coefficient. Beneath the scatterplot is the residual plot.

2. a. Describe the shape of the distribution of fruit mass versus fruit circumference with respect to shape. Does the distribution surprise you given what you saw in the spreadsheet? Explain your thinking.

   b. Describe the shape of the residual plot of fruit mass versus fruit circumference.

   c. What is the correlation coefficient? What does that suggest about the linear relationship between fruit circumference and fruit mass?

d. Would it be appropriate to use this least-squares regression line to make a prediction about the mass of a fruit based on its circumference? Explain?

The arrow will allow you to explore several transformations of the fruit data. For each transformation, you will see the scatterplot and residual plot of the transformed data, the equation of the least-squares regression line, and the correlation coefficient. The goal is to find a transformation that changes numbers in ways that makes the graph more linear. Using the arrow will display square root, semi-log, and log-log transformations.

3. Click on the arrow in the upper panel to take the square root of all of the fruit masses.
   a. How do the scatterplot and residual plot change with respect to shape? What is the new correlation coefficient?

   b. How would you write the equation for the least-squares regression line for the transformed data? (Hint: Look at the variable names on each axis to see where you need to include "sqrt".)

   c. Do the transformed data have a linear relationship? Explain why or why not.

4. Another way to change the shape of the distribution is to take the common (base 10) or natural (base e) logarithm. Click on the right arrow again to take the common logarithm of the fruit mass values.
   a. How do the scatterplot and residual plot change with respect to shape? What is the new correlation coefficient?

   b. Did this logarithmic transformation “fix” the problem of a non-linear relationship between fruit circumference and fruit mass? Explain your reasoning.

5. a. Click on the right arrow again, and describe the new transformation.

   b. How do the scatterplot and residual plot change with respect to shape? What is the new correlation coefficient?

c. Do the transformed data have a linear relationship? Explain.

6. Click on the right arrow again.
   a. This is a log-log transformation used to check for a power relationship in the original data. Explain why the name makes sense.

   b. How do the scatterplot and residual plot change with respect to shape? What is the new correlation coefficient? Did the log-log transformation “fix” the curve in the original data so that it would be possible to use this model for prediction – that is, do the transformed data have a linear relationship?

   c. How would you write the equation for the least-squares regression line for the transformed data?

   d. Use your equation to predict the mass of a spherical watermelon with a circumference of 44 cm and one with a circumference of 68 cm. Which prediction do you feel is more reliable, and why?

Move to page 2.1.

7. The data set on this page shows the lengths in inches and the weights in pounds for 25 alligators. What do you think a scatterplot of these bivariate data will look like?

Move to page 2.2.

Page 2.2 shows a scatterplot of the alligator weight vs. length with the least-squares regression line. The scatterplot on the lower screen is the residual plot.

8. Describe the distribution of alligator weight versus alligator length shown in the scatterplot with respect to shape. Does the distribution surprise you given what you saw in the spreadsheet?

9. a. Use the arrow in the upper panel to check four different transformations of the alligator data. Which transformation made the scatterplot of the data the most linear and the residual plot the most random? Explain your reasoning.

   b. How would you write the equation for the least-squares line for the transformation that seems best at eliminating the curvature in the alligator data?

   c. Predict the weight of an alligator whose length is 140 cm.

Data used in this activity:

Concept of using fruit (name of fruit, circumference in cm, mass in grams)--thanks to Tim Erickson of

http://www.eeps.com/zoo/index.html

Alligator data (length in inches and weight in pounds)--thanks to Richard Scheaffer