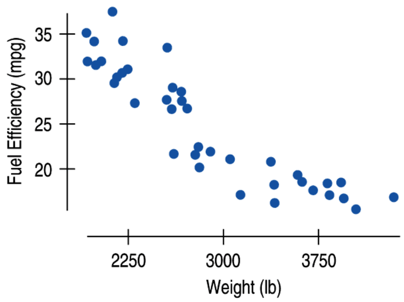
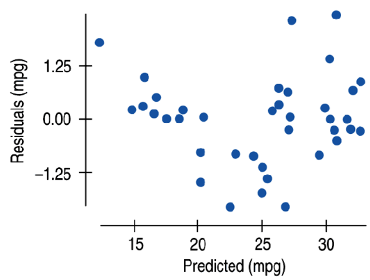
Chapter 10: Re-expressing Data: Get It Straight!

Re-expressing Data

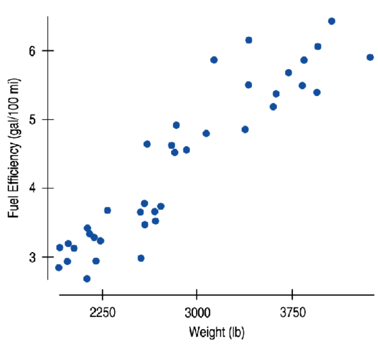
We cannot use a linear model unless the relationship between the two variables is ________________. If the relationship is nonlinear (which we can verify by examining the ___________________________) we can try ________________ the data. Then we can fit and use a simple linear model. To re-express the data, we perform some mathematical operation on the data values such as taking the ________________ , taking the ________________ , or taking the ________________ .

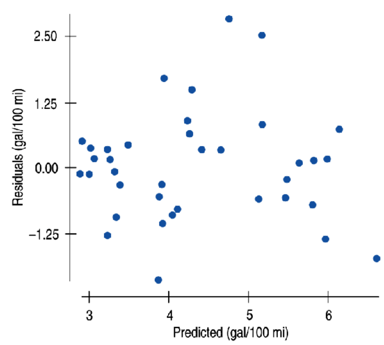
For example, consider the relationship between the weight of cars and their fuel efficiency (miles per gallon). What do the scatterplot and residual plots reveal? ________________________________ ________________





If we take the reciprocal of the *y*-values, we get the following scatterplot and residual plot. What do these plots reveal? ____





There are several reasons we may want to re-express our data:

1. To make the distribution of a variable more ________________ .
2. To make the ________________ of several groups more alike.
3. To make the form of a scatterplot more ________________ .
4. To make the scatter in a scatterplot more ________________ ________________ .

Re-expressing Data Using Logarithms

An equation of the form $y = a + bx$ is used to model ________________ data.

The process of transforming nonlinear data into linear data is called ________________. In order to linearize certain types of data we use properties of ____________________.

PROPERTIES OF LOGARITHMS:

1. $\log ab =$
2. $\log \frac{a}{b} =$
3. $\log x^p =$

Case 1: Consider the following set of Linear Data representing an account balance as a function of time:

x: time (months)	0	48	96	144	192	240
y: account balance ($)	100	580	1060	1540	2020	2500

Describe the pattern of change…

The relationship between *x* and *y* is __________ if, for equal increments of *x*, we __________ a fixed increment to *y*.

Case 2: Consider the following set of Nonlinear Data representing an account balance as a function of time:

x: time (months)	0	48	96	144	192	240
y: account balance ($)	100	161.22	259.93	419.06	675.62	1089.30

Describe the pattern of change…

The relationship between *x* and *y* is ____________________ if, for equal increments of *x*, we _______________ a fixed increment by *y*. This increment is called the ____________________.

We want to find the best fitting model to represent our data.

- For the linear data, we use least-squares regression to find the best fitting _____________.
- For the nonlinear data, the best fitting model would be an exponential ________________.

PROBLEM: We cannot use least-squares regression for the nonlinear data because least-squares regression depends upon correlation, which only measures the strength of _______________ relationships.

SOLUTION: We transform the *nonlinear data* into *linear data*, and then use least-squares regression to determine the best fitting __________for the transformed data.

Finally, do a _______________ transformation to turn the linear equation back into a nonlinear equation which will model our original *nonlinear data*.

Linearizing Exponential Functions:

(We want to write an exponential function of the form $y = a \cdot b^x$ as a function of the form $y = a + bx$).

$y = a \cdot b^x$ (_____ , _____ are variables and _____ , _____ are constants)

This is in the general form ________________, which is linear.

So, the graph of (var1, var2) is linear. This means the graph of $(x, \log y)$ is linear.

CONCLUSIONS:

1. If the graph of ________________ is linear, then the graph of ________________ is exponential.

2. If the graph of ________________ is exponential, then the graph of ________________ is linear.

Once we have linearized our data, we can use least-squares regression on the transformed data $(x, \log y)$ to find the best fitting <u>linear</u> model.

PRACTICE:

Linearize the data for Case 2 and find the least-squares regression line for the transformed data.

Then, do a reverse transformation to turn the linear equation back into an exponential equation.

Compare this to the equation the calculator gives when performing exponential regression on the Case 2 data.

Linearizing Power Functions:

(We want to write a power function of the form $y = ax^b$ as a function of the form $y = a + bx$).

$y = ax^b$ (_____ , _____ are variables and _____ , _____ are constants)

This is in the general form ________________, which is linear.

So, the graph of (var1, var2) is linear. This means the graph of $(\log x, \log y)$ is linear.

Case 3: Consider the following set of Nonlinear Data representing the average length and weight at different ages for Atlantic Ocean rockfish:

x: age (years)	0	4	8	12	16	20
y: weight (grams)	0	48	192	432	768	1200

PRACTICE:

Linearize the data for Case 3 and find the least-squares regression line for the transformed data.

Then, do a reverse transformation to turn the linear equation back into a power equation.

Compare this to the equation the calculator gives when performing power regression on the Case 3 data.