

Ch 12.2 Transforming Linearity

Applying a function such as the logarithm or square root to a quantitative variable is called **transforming** the data. We will see in this section that understanding how simple functions work helps us choose and use transformations to straighten nonlinear patterns.

Example 6 Health and Wealth
Straightening out a curved pattern

The Gapminder Web site, www.gapminder.org, provides loads of data on the health and well-being of the world's inhabitants. Figure 12.9 on the next page is a scatterplot of data from Gapminder.¹⁴ The individuals are all the world's nations for which data are available. The explanatory variable is a measure of how rich a country is: income per person. The response variable is life expectancy at birth. We expect people in richer countries to live longer because they have better access to medical care and typically lead healthier lives. The overall pattern of the scatterplot does show this, but the relationship is not linear. Life expectancy rises very quickly as income per person increases and then levels off. People in very rich countries such as the United States live no longer than people in poorer but not extremely poor nations. In some less wealthy countries, people live longer than in the United States.

Figure 12.10 Scatterplot of life expectancy against income per person (on a logarithmic scale) for many nations.

Mar 7-12:38 PM

Example 6 Go Fish!
Transforming with powers

Imagine that you have been put in charge of organizing a fishing tournament in which prizes will be given for the heaviest Atlantic Ocean rockfish caught. You know that many of the fish caught during the tournament will be measured and released. You are also aware that using delicate scales to try to weigh a fish that is flopping around in a moving boat will probably not yield very accurate results. It would be much easier to measure the length of the fish while on the boat. What you need is a way to convert the length of the fish to its weight.

You contact the nearby marine research laboratory, and they provide reference data (on the length (in centimeters) and weight (in grams)) for Atlantic Ocean rockfish of several sizes.¹⁵

Length	5.2	8.5	11.5	14.3	16.8	19.2	21.3	23.3	25.0	26.7
Weight	2	8	21	38	69	117	148	190	264	283

Figure 12.11 is a scatterplot of the data. Note the clear curved shape.

This transformation of the explanatory variable helps us produce a graph that is quite linear:

Figure 12.12 The scatterplot of $\sqrt[3]{\text{weight}}$ versus length is linear.

Mar 7-12:45 PM

Example 7 Go Fish!
Transforming with Powers and Roots

Here is Minitab output from separate regression analyses of the two sets of transformed Atlantic Ocean rockfish data.

(a) Give the equation of the least-squares regression line. Define any variables you use.

(b) Suppose a contestant in the fishing tournament catches an Atlantic Ocean rockfish that's 30 centimeters long. Use the model from part (a) to predict the fish's weight. Show your work.

(a) Transformation 1: $\text{weight} = 4.066 + 0.0146774(\text{length}^3)$

Predictor	Coef	SE Coef	T	P
Constant	4.066	6.902	0.59	0.5
Length^3	0.0146774	0.0002404	61.07	0.000

S = 18.8412 R-Sq = 99.5% R-Sq(adj) = 99.5%

(b) Transformation 2: $\text{weight} = -0.02204 + 0.246616(\text{length})$

Predictor	Coef	SE Coef	T	P
Constant	-0.02204	0.07762	-0.28	0.780
Length	0.246616	0.002868	86.00	0.000

S = 0.124161 R-Sq = 99.8% R-Sq(adj) = 99.7%

Mar 7-12:48 PM

Health and wealth of nations (Scatter Plot 12)

The regression line is

$$\text{predicted life expectancy} = 19.5 + 13.2 \log(\text{income})$$

predicted life expectancy = $19.5 + 13.2 \log(42,296.20) = 80.567$ years

income per person of \$42,296.20

Mar 7-12:51 PM

Power Model vs Exponential Model

Power model

(log x, log y) or

(log explanatory variable, response variable)

If the scatterplot of logarithm (or natural logarithm) of the response variable values and the logarithm (or natural logarithm) of the explanatory values has a linear form, then the 2 variables can be modeled using a power function.

Exponential Model

(x, log y) or

(explanatory x, log response variable)

If the scatterplot of logarithm (or natural logarithm) of the response variable values and the original explanatory values has a linear form, then the 2 variables can be modeled using an exponential function.

Regression Analysis: log(Weight) versus log(Length)

Predictor	Coef	SE Coef	T	P
Constant	-1.89940	0.03799	-49.99	0.000
log(Length)	3.04942	0.02764	110.31	0.000

S = 0.0281823 R-Sq = 99.9% R-Sq(adj) = 99.8%

(a) Based on the output, explain why it would be reasonable to use a power model to describe the relationship between weight and length for Atlantic Ocean rockfish. **Power Model - taking logs of both variables.**

(b) Give the equation of the least-squares regression line. Be sure to define any variables you use.

$$\log(\text{weight}) = -1.89940 + 3.04942 \log(\text{length})$$

Example 9 Go Fish!
Fishing predictions

PROBLEM: Suppose a contestant in the fishing tournament catches an Atlantic Ocean rockfish that's 36 centimeters long. Use the model from part (b) of the previous example to predict the fish's weight. Show your work.

SOLUTION: For a length of 36 centimeters, we have

$$\log(\text{weight}) = -1.89940 + 3.04942 \log(36) = 2.8464$$

To find the predicted weight, we use the definition of a logarithm as an exponent:

$$\log_{10}(\text{weight}) = 2.8464$$

$$\text{weight} = 10^{2.8464} \approx 702.1$$

This model predicts that a 36-centimeter-long rockfish will weigh about 702 grams.

log base Answer = exponent
ln is base e

Mar 7-12:55 PM

Mar 7-12:53 PM

Example 11 Moore's Law and Computer Chips

Exponential transformations and exponential models

Gordon Moore, one of the founders of Intel Corporation, predicted in 1975 that the number of transistors on an integrated circuit chip would double every 18 months. This is Moore's law, one way to measure the revolution in computing: more are data and the cubic performance of transistors for fast microprocessors.

Year	Transistors
1971	2,300
1974	2,900
1976	2,900
1978	29,000
1982	100,000
1985	275,000
1989	1,200,000
1993	3,100,000
1995	7,500,000
1997	15,000,000
1999	40,000,000
2001	90,000,000
2003	200,000,000
2004	290,000,000
2006	1,000,000,000
2008	1,900,000,000
2010	2,300,000,000

(x, ln y)

ln(transistors) = 7.0647 + 0.3668(ln years since 1970)

ln(transistors) = 7.0647 + 0.3668(50) = 25.562

To find the predicted number of transistors, we use the definition of a logarithm as is expected.

ln(transistors) = 25.562 ⇒ ln(transistors) = 25.562

transistors = $e^{25.562} \approx 1.028 \cdot 10^{11}$

Mar 7-1:08 PM

Day 2: Putting It All Together: Which Transformation Should We Choose?

Enter the values of the explanatory variable in L1/list1 and the values of the response variable in L2/list2. Make a scatterplot of y versus x and confirm that there is a curved pattern.

Example 12 What's a Planet, Anyway?

Power models and logarithm transformations

On July 31, 2005, a team of astronomers announced that they had discovered what appeared to be a new planet in our solar system. They had first observed this object almost two years earlier using a telescope at Caltech's Palomar Observatory in California. Originally named U8313, the potential planet is bigger than Pluto and has an average distance of about 9.5 billion miles from the sun. (For reference, Earth is about 93 million miles from the sun.) Could this new astronomical body, now called Eris, be a new planet?

At the time of the discovery, there were nine known planets in our solar system. Here are data on the distance from the sun and period of revolution of those planets. Note that distance is measured in astronomical units (AU), the number of Earth distances the object is from the sun.¹⁷

Planet	Distance from sun (astronomical units)	Period of revolution (Earth years)
Mercury	0.387	0.241
Venus	0.723	0.615
Earth	1.000	1.000
Mars	1.524	1.881
Jupiter	5.203	11.862
Saturn	9.539	29.456
Uranus	19.191	84.070
Neptune	30.061	164.810
Pluto	39.529	248.530

Mar 8-12:07 PM

Figure 12.18 is a scatterplot of the planetary data. This appears to be a strong curved relationship between distance from the sun and period of revolution.

ln(distance) stores in another list
ln(period) stores in another list
create scatter plots of different sets of data to see if it **looks more** exponential.

Figure 12.19 (a) A scatterplot of ln(period) versus distance. (b) A scatterplot of ln(period) versus ln(distance).

(a) The scatterplot of ln(period) versus distance is clearly curved, so an exponential model would not be appropriate. However, the graph of ln(period) versus ln(distance) has a strong linear pattern, indicating that a power model would be more appropriate.

(b) Eris's average distance from the sun is 102.15 AU. Using the value for distance in our model from part (b) gives

$\ln(\text{period}) = 0.002544 + 1.49996 \ln(102.15) = 6.859$

To predict the period, we have to undo the logarithm transformation:

$\ln(\text{period}) = 6.859 \Rightarrow \text{period} = e^{6.859} \approx 1052 \text{ years}$

We wouldn't want to wait for Eris to make a full revolution to see if our prediction is accurate!

Mar 8-10:27 AM

Construct a residual plot to look for any departures from the linear pattern. For Xlist, enter the list you used as the explanatory variable in the linear regression calculation. For Ylist, use the RESID list stored in the calculator. For the planet data, we used L3/list3 as the Xlist.

(d) Eris's value for ln(distance) is ln(102.15) = 4.626, which would fall at the far right of the residual plot, where all the residuals are positive. Because residual = actual y - predicted y seems likely to be positive, we would expect our prediction to be too low.

To make a prediction for a specific value of the explanatory variable, compute log x or ln x, as appropriate. Then use Y1(x) to obtain the predicted value of log y or ln y. To get the predicted value of y, use 10^Ans or e^Ans to undo the logarithm transformation. Here's our prediction of the period of revolution for Eris, which is at a distance of 102.15 AU from the sun:

Y1(102.15)	4.626442321
Y1(Ans)	6.939274784
e^Y1(Ans)	1052.621505

Mar 8-12:12 PM

CHECK YOUR UNDERSTANDING

One said that about the only thing that's true about life is that we'll all die sometime. Many adults plan ahead for their eventual departure by purchasing life insurance. Many different types of life insurance policies are available. Some provide coverage throughout an individual's life (whole life), while others last only for a specified number of years (term life). The table below shows monthly premiums for a 30-year term life insurance policy in dollars for various ages.

Age (years)	Monthly premium
40	\$29
45	\$46
50	\$68
55	\$106
60	\$157
65	\$257

1. Use each model to predict how much a 30-year-old would pay for such a policy. Show your work.

2. What type of function—linear, power, or exponential—best describes the relationship between age and monthly premium? Explain.

Correct Answer:
Linear Answer:
 $\text{premium} = 1.171(\text{age}) - 341 = 1.171(30) = \171.54 (approx.)
Exponential Answer:
 $\ln(\text{premium}) = -12.85 + 4.416(\ln \text{age}) = 4.9500$, $e^{4.9500} = \$141.50$ (approx.)
Power Answer:
 $\ln(\text{premium}) = -0.0961 + 0.0078(\ln \text{age}) = 4.9502$, $e^{4.9502} = \$141.50$ (approx.)

Correct Answer:
Exponential Answer: 2, because the scatterplot showing ln(premium) versus age was the most linear and the model had the most randomly scattered residuals.

Mar 8-3:25 PM