SUPPLEMENTAL MATERIAL: TRANSFORMING DATA, DECISION ANALYSIS

Introduction

This brief supplement presents material useful to some readers of *IPS* that was removed from the main text for reasons of length. It has two sections:

1. A section on *transforming relationships.* This is intended to follow Section 2.6 of *IPS,* which now ends Chapter 2.

2. A supplement to Section 4.5, briefly discussing *decision analysis.* The natural place for this material is immediately following the discussion of tree diagrams, *IPS* page 301.

Supplemental Section 2.6

Transforming Relationships

Let's start by revisiting Example 2.18 on page 120 of *IPS*. How is the weight of an animal's brain related to the weight of its body? Figure 2.31 (which repeats Figure 2.17) is a scatterplot of brain weight against body weight for 96 species of mammals.[1] The line is the least-squares regression line for predicting brain weight from body weight. The outliers are interesting. We might say that dolphins and humans are smart, hippos are dumb, and African elephants are just big. That's because dolphins and humans have larger brains than their body weights suggest, hippos have smaller brains, and the elephant is much heavier than any other mammal in both body and brain.

EXAMPLE 2.37 The plot in Figure 2.31 is not very satisfactory. Most mammals are so small relative to elephants and hippos that their points overlap to form a blob in the lower-left corner of the plot. The correlation between brain weight and body weight is $r = 0.86$, but this is misleading. If we remove the elephant, the correlation for the other 95 species is $r = 0.50$. Figure 2.32 is a scatterplot of the data with the four outliers removed to allow a closer look at the other 92 observations. We can now see that the relationship is not linear. It bends to the right as body weight increases.

Biologists know that data on sizes often behave better if we take *logarithms*

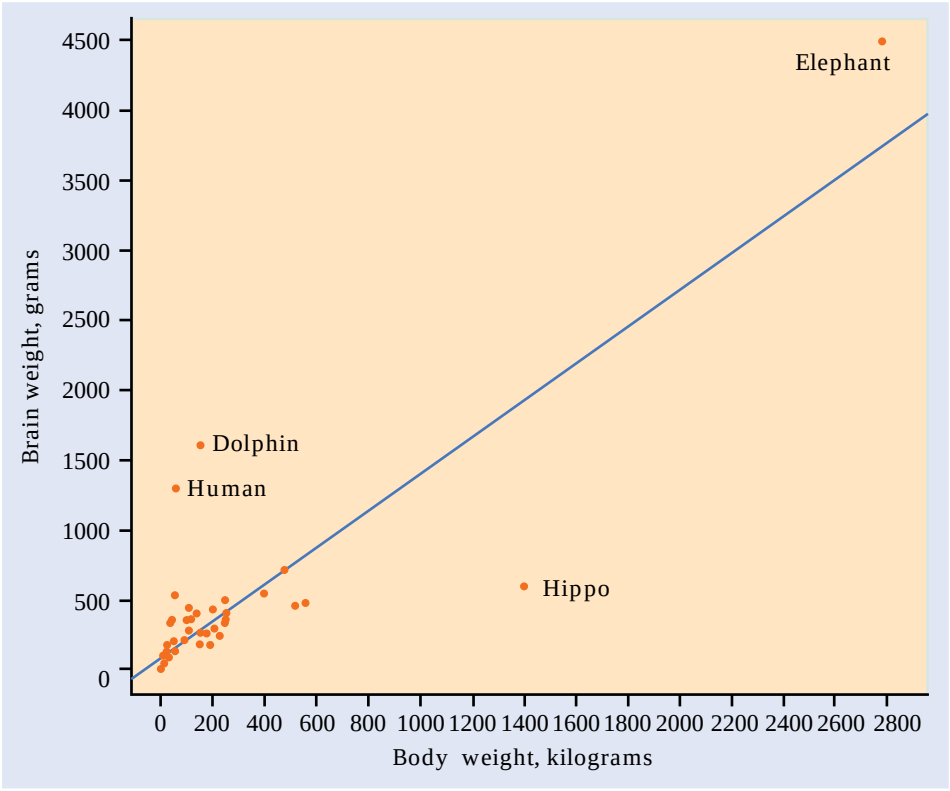


FIGURE 2.31 Scatterplot of brain weight against body weight for 96 species of mammals.

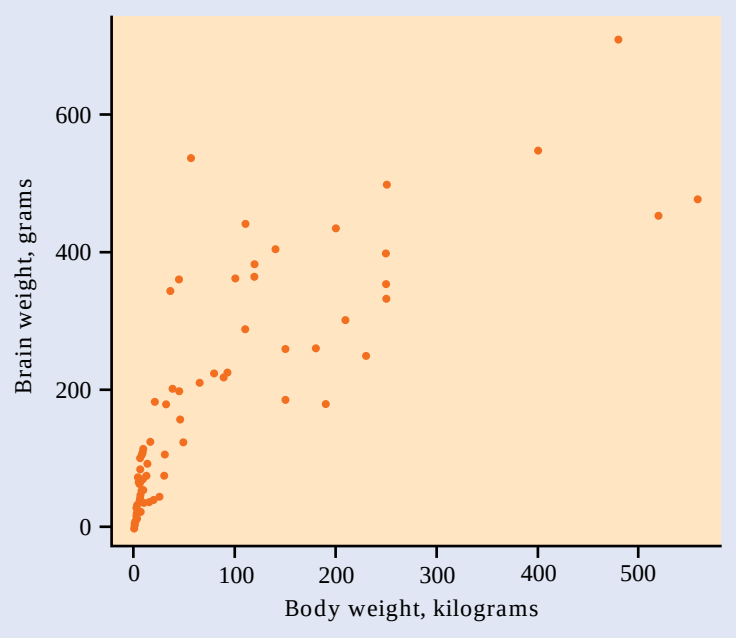


FIGURE 2.32 Scatterplot of brain weight against body weight for mammals, with outliers removed.

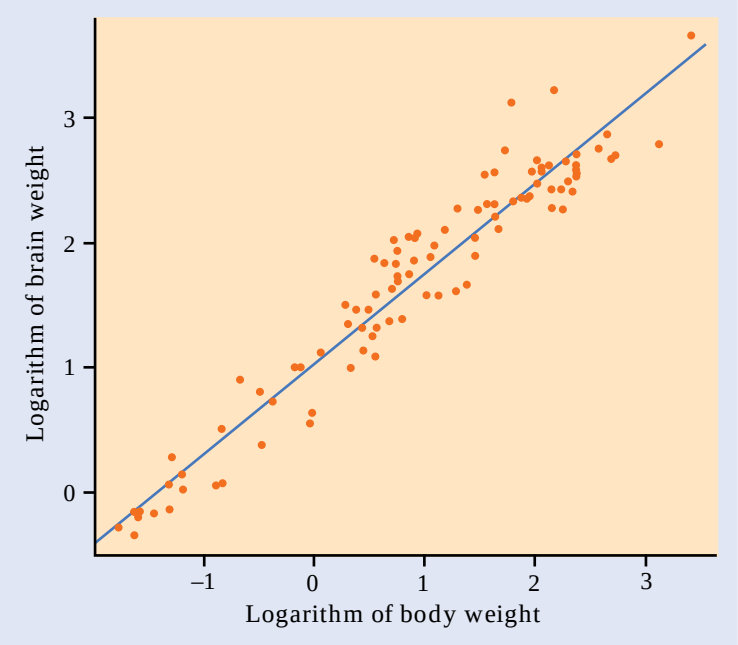


FIGURE 2.33 Scatterplot of the logarithm of brain weight against the logarithm of bodyweight for 96 species of mammals.

before doing more analysis. Figure 2.33 plots the logarithm of brain weight against the logarithm of body weight for all 96 species. The effect is almost magical. There are no longer any extreme outliers or very influential observations. The pattern is very linear, with correlation $r = 0.96$. The vertical spread about the least-squares line is similar everywhere, so that predictions of brain weight from body weight will be about equally precise for any body weight (in the log scale).

Example 2.37 shows that working with a *function* of our original measurements can greatly simplify statistical analysis. Applying a function such as the logarithm or square root to a quantitative variable is called **transforming** or **reexpressing** the data. We will see in this section that understanding how simple functions work helps us choose and use transformations. Because we may want to transform either the explanatory variable x or the response variable y in a scatterplot, or both, we will call the variable t when talking about transforming in general.

transforming
reexpressing

First steps in transforming

Transforming data amounts to changing the scale of measurement that was used when the data were collected. We can choose to measure temperature in degrees Fahrenheit or in degrees Celsius, distance in miles or in kilometers. These changes of units are *linear transformations.* **Linear transformations cannot straighten a curved relationship between two variables.** To do that, we resort to functions that are

not linear. The logarithm, applied in Example 2.37, is a nonlinear function. Here are some others.

- How shall we measure the size of a sphere or of such roughly spherical objects as grains of sand or bubbles in a liquid? The size of a sphere can be expressed in terms of the diameter t, in terms of surface area (proportional to t^2), or in terms of volume (proportional to t^3). Any one of these *powers* of the diameter may be natural in a particular application.

- We commonly measure the fuel consumption of a car in miles per gallon—that is, how many miles the car travels on 1 gallon of fuel. Engineers prefer to measure in gallons per mile—that is, how many gallons of fuel the car needs to travel 1 mile. This is a *reciprocal* transformation. A car that gets 25 miles per gallon uses

$$\frac{1}{\text{miles per gallon}} = \frac{1}{25} = 0.04 \text{ gallons per mile}$$

  The reciprocal is a *negative power* $1/t = t^{-1}$.

The transformations we have mentioned—linear, positive and negative powers, and logarithms—are those used in most statistical problems. They are all *monotonic.*

Monotonic Functions

A **monotonic function** $f(t)$ moves in one direction as its argument t increases.

A **monotonic increasing** function preserves the order of data. That is, if $a > b$, then $f(a) > f(b)$.

A **monotonic decreasing** function reverses the order of data. That is, if $a > b$, then $f(a) < f(b)$.

The graph of a linear function is a straight line. The graph of a monotonic increasing function is increasing everywhere. A monotonic decreasing function has a graph that is decreasing everywhere. A function can be monotonic over some range of t without being everywhere monotonic. For example, the square function t^2 is monotonic increasing for $t \geq 0$. If the range of t includes both positive and negative values, the square is not monotonic—it decreases as t increases for negative values of t and increases as t increases for positive values.

Figure 2.34 compares three monotonic increasing functions and three monotonic decreasing functions for positive values of the argument t. Many variables take only 0 or positive values, so we are particularly interested in how functions behave for positive values of t. The increasing functions are

Linear	$a + bt$, slope $b > 0$
Square	t^2
Logarithm	$\log t$

The decreasing functions in the lower panel of Figure 2.34 are

Linear	$a + bt$, slope $b < 0$
Reciprocal square root	$1/\sqrt{t}$, or $t^{-1/2}$
Reciprocal	$1/t$, or t^{-1}

Nonlinear monotonic transformations change data enough to alter the shape of distributions and the form of relations between two variables, yet are simple enough to preserve order and allow recovery of the original data. We will concentrate on powers and logarithms. The even-numbered powers t^2, t^4, and so on are monotonic increasing for $t \geq 0$, but not when t can take both negative and positive values. The logarithm is not even defined unless $t > 0$. Our strategy for transforming data is therefore as follows:

1. If the variable to be transformed takes values that are 0 or negative, first apply a linear transformation to make the values all positive. Often we just add a constant to all the observations.

2. Then choose a power or logarithmic transformation that simplifies the data, for example, one that approximately straightens a scatterplot.

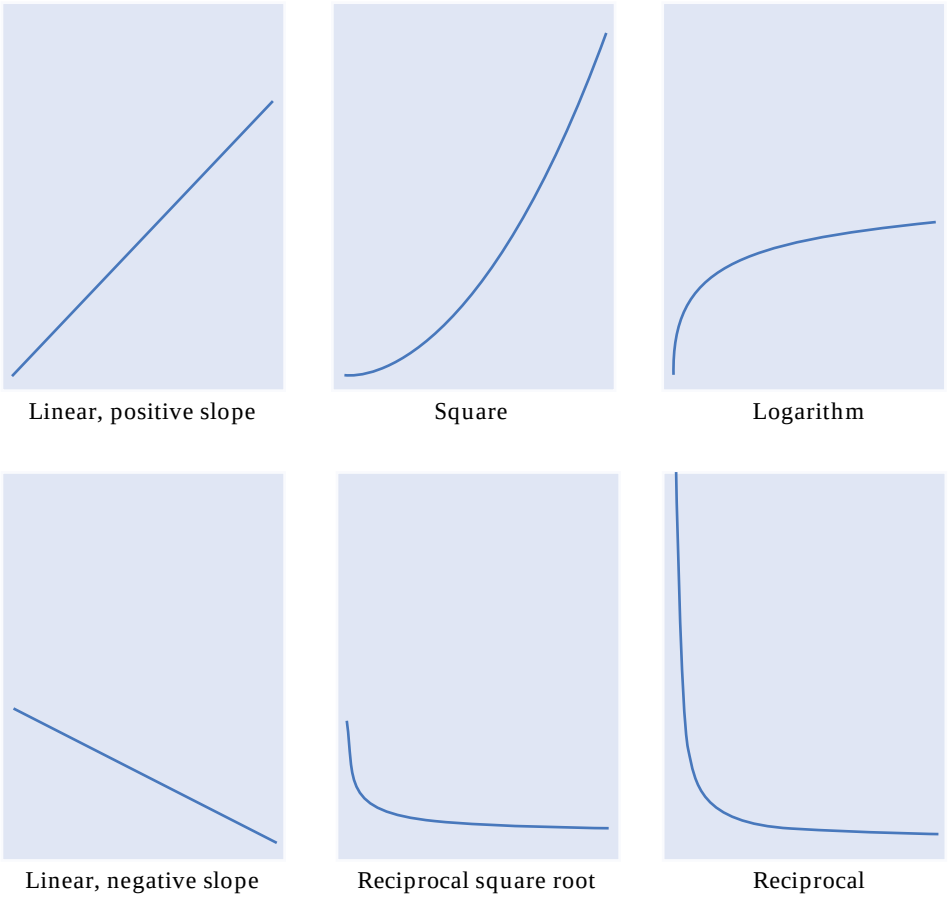


FIGURE 2.34 Monotonic functions, increasing (top panel) and decreasing (bottom panel).

The ladder of power transformations

Though simple in algebraic form and easy to compute with a calculator, the power and logarithm functions are varied in their behavior. It is natural to think of powers such as

$$\ldots, t^{-1}, t^{-1/2}, t^{1/2}, t, t^2, \ldots$$

as a hierarchy or ladder. Some facts about this ladder will help us choose transformations. In all cases, we look only at positive values of the argument t.

Monotonicity of Power Functions

Power functions t^p for positive powers p are monotonic increasing for values $t > 0$. They preserve the order of observations. This is also true of the logarithm.

Power functions for negative powers p are monotonic decreasing for values $t > 0$. They reverse the order of the observations.

It is hard to interpret graphs when the order of the original observations has been reversed. We can make a negative power such as the reciprocal $1/t$ monotonic increasing rather than monotonic decreasing by using $-1/t$ instead. Figure 2.35 takes this idea a step farther. This graph compares the ladder of power functions in the form

$$\frac{t^p - 1}{p}$$

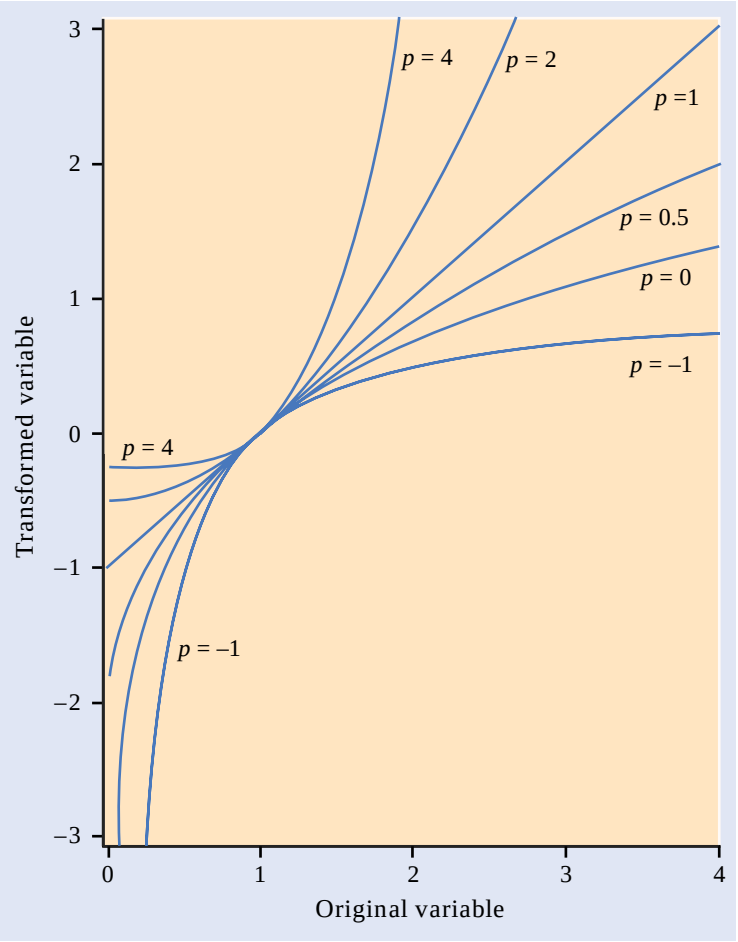


FIGURE 2.35 The ladder of power and logarithm transformations. For easier comparison, the functions have been scaled so that all meet at $t = 1$ and have slope 1 at that point.

The reciprocal (power $p = -1$), for example, is graphed as

$$\frac{1/t - 1}{-1} = 1 - \frac{1}{t}$$

This linear transformation does not change the nature of the power functions t^p, except that all are now monotonic increasing. It is chosen so that every power has the value 0 at $t = 1$ and also has slope 1 at that point. So the graphs in Figure 2.35 all touch at $t = 1$ and go through that point at the same slope.

Look at the $p = 0$ graph in Figure 2.35. The 0th power t^0 is just the constant 1, which is not very useful. The $p = 0$ entry in the figure is not constant. In fact, it is the logarithm, $\log t$. That is, **the logarithm fits into the ladder of power transformations at $p = 0$.**[2]

Figure 2.35 displays another key fact about these functions. The graph of a linear function (power $p = 1$) is a straight line. Powers greater than 1 give graphs that bend upward. That is, the transformed variable grows ever faster as t gets larger. Powers less than 1 give graphs that bend downward. The transformed values continue to grow with t, but at a rate that decreases as t increases. What is more, the sharpness of the bend increases as we move away from $p = 1$ in either direction.

Concavity of Power Functions

Power transformations t^p for powers p greater than 1 are *concave up;* that is, they have the shape $\bigcup$. These transformations push out the right tail of a distribution and pull in the left tail. This effect gets stronger as the power p moves up away from 1.

Power transformations t^p for powers p less than 1 (and the logarithm for $p = 0$) are *concave down;* that is, they have the shape $\bigcap$. These transformations pull in the right tail of a distribution and push out the left tail. This effect gets stronger as the power p moves down away from 1.

EXAMPLE 2.38 Figure 2.36(a) is a scatterplot of data from the World Bank.[3] The individuals are all the world's nations for which data are available. The explanatory variable x is a measure of how rich a country is: the gross domestic product (GDP) per person. GDP is the total value of the goods and services produced in a country, converted into dollars. The response variable y is life expectancy at birth.

Life expectancy increases in richer nations, but only up to a point. The pattern in Figure 2.36(a) at first rises rapidly as GDP increases but then levels out. Three African nations (Botswana, Gabon, and Namibia) are outliers with much lower life expectancy than the overall pattern suggests. Can we straighten the overall pattern by transforming?

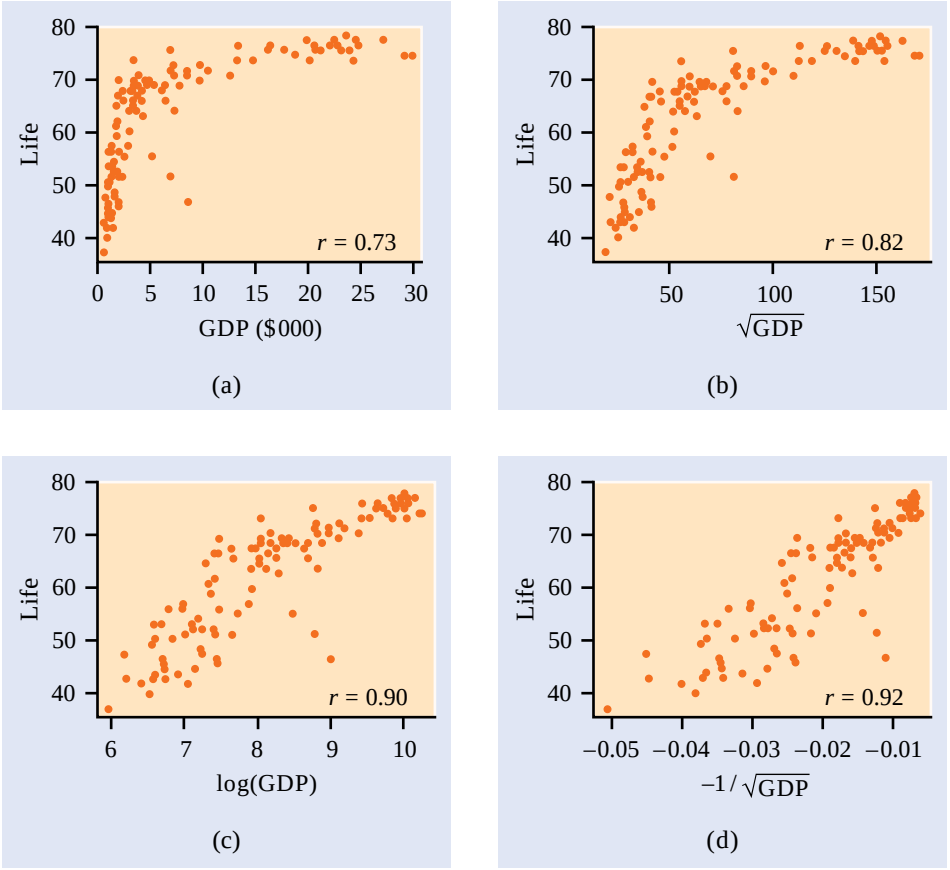


FIGURE 2.36 The ladder of transformations at work. The data are life expectancy and gross domestic product (GDP) for 115 nations. Panel (a) displays the original data. Panels (b), (c), and (d) transform GDP, moving down the ladder away from linear functions.

Life expectancy does not have a large range, but we can see that the distribution of GDP is right-skewed and very spread out. So GDP is a better candidate for transformation. We want to pull in the long right tail, so we try transformations with $p < 1$. Figures 2.36(b), (c), and (d) show the results of three transformations of GDP. The r-value in each figure is the correlation when the three outliers are omitted.

The square root $\sqrt{x}$, with $p = 1/2$, reduces the curvature of the scatterplot, but not enough. The logarithm $\log x$ ($p = 0$) straightens the pattern more, but it still bends to the right. The reciprocal square root $1/\sqrt{x}$, with $p = -1/2$, gives a pattern that is quite straight except for the outliers. To avoid reversing the order of the observations, we actually used $-1/\sqrt{x}$.

Example 2.38 shows the ladder of powers at work. As we move down the ladder from linear transformations (power $p = 1$), the scatterplot gets straighter. Moving farther down the ladder, to the reciprocal $1/x = x^{-1}$, begins to bend the plot in the other direction. But this "try it and see" approach isn't very satisfactory. That life expectancy depends linearly on $1/\sqrt{\text{GDP}}$ does not increase our understanding of the relationship between the health and wealth of nations. We don't recommend just pushing buttons on your calculator to try to straighten a scatterplot.

It is much more satisfactory to begin with a theory or mathematical model that we expect to describe a relationship. The transformation needed to make the rela-

tionship linear is then a consequence of the model. One of the most common models is *exponential growth.*

Exponential growth

A variable grows linearly over time if it *adds* a fixed increment in each equal time period. Exponential growth occurs when a variable is *multiplied* by a fixed number in each time period. To grasp the effect of multiplicative growth, consider a population of bacteria in which each bacterium splits into two each hour. Beginning with a single bacterium, we have 2 after one hour, 4 at the end of two hours, 8 after three hours, then 16, 32, 64, 128, and so on. These first few numbers are deceiving. After 1 day of doubling each hour, there are 2^{24} (16,777,216) bacteria in the population. That number then doubles the next hour! Try successive multiplications by 2 on your calculator to see for yourself the very rapid increase after a slow start. Figure 2.37 shows the growth of the bacteria population over 24 hours. For the first 15 hours, the population is too small to rise visibly above the zero level on the graph. It is characteristic of exponential growth that the increase appears slow for a long period, then seems to explode.

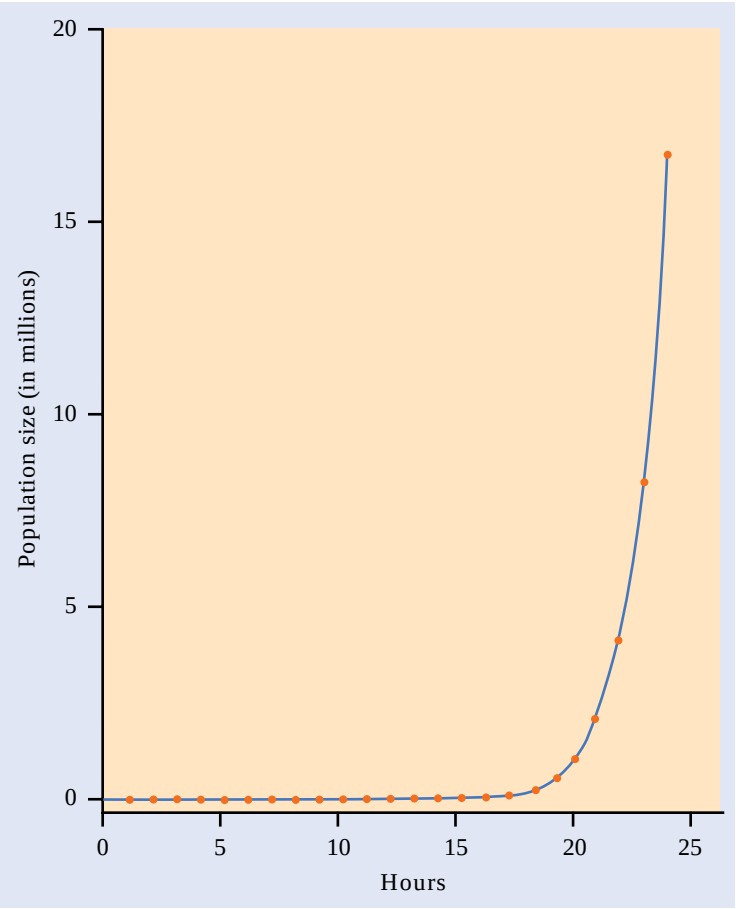


FIGURE 2.37 Growth of a bacteria population that doubles each hour.

Linear versus Exponential Growth

Linear growth increases by a fixed *amount* in each equal time period. **Exponential growth** increases by a fixed *percentage* of the previous total.

Populations of living things—like bacteria—tend to grow exponentially if not restrained by outside limits such as lack of food or space. More pleasantly, money also displays exponential growth when returns to an investment are compounded. Compounding means that last period's income earns income this period.

EXAMPLE 2.39 A dollar invested at an annual rate of 6% turns into \$1.06 in a year. The original dollar remains and has earned \$0.06 in interest. That is, 6% annual interest means that any amount on deposit for the entire year is multiplied by 1.06. If the \$1.06 remains invested for a second year, the new amount is therefore 1.06×1.06, or 1.06^2. That is only \$1.12, but this in turn is multiplied by 1.06 during the third year, and so on. After x years, the dollar has become 1.06^x dollars.

If the Native Americans who sold Manhattan Island for \$24 in 1626 had deposited the \$24 in a savings account at 6% annual interest, they would now have almost \$80 billion. Our savings accounts don't make us billionaires, because we don't stay around long enough. A century of growth at 6% per year turns \$24 into \$8143. That's 1.06^{100} times \$24. By 1826, two

centuries after the sale, the account would hold a bit over \$2.7 million. Only after a patient 302 years do we finally reach \$1 billion. That's real money, but 302 years is a long time.

The count of bacteria after x hours is 2^x. The value of \$24 invested for x years at 6% interest is 24×1.06^x. Both are examples of the **exponential growth model** $y = a \times b^x$ for different constants a and b. In this model, the response y is multiplied by b in each time period.

exponential growth model

EXAMPLE 2.40 Does the exponential growth model sometimes describe real data that don't arise from any obvious process of multiplying by a fixed number over and over again? Let's look at the world's production of oil. Petroleum is the most important source of energy for the developed nations and has been the cause of economic dislocation and even war. Table 2.10 and Figure 2.38 show the growth of annual world crude oil production, measured in millions of barrels per year.[4] There is an increasing trend, but the overall pattern is not linear. Oil production has increased much faster than linear growth. The pattern of growth follows a smooth curve until 1973, when a Mideast war touched off a vast price increase and a change in the previous pattern of production. In fact, the curve of oil production from 1880 to 1973 looks a lot like an exponential curve. Is this exponential growth?

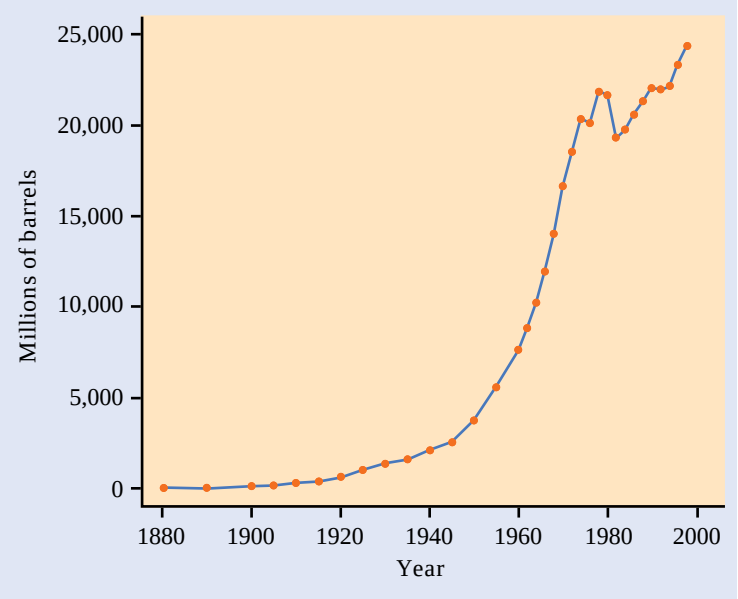


FIGURE 2.38 World production of crude oil from 1880 to 1998. This time plot has a regular pattern prior to 1973.

TABLE 2.10

Annual world crude oil production, 1880–1998 (millions of barrels)

Year	Mbbl.	Year	Mbbl.	Year	Mbbl.
1880	30	1945	2,595	1976	20,188
1890	77	1950	3,803	1978	21,922
1900	149	1955	5,626	1980	21,722
1905	215	1960	7,674	1982	19,411
1910	328	1962	8,882	1984	19,837
1915	432	1964	10,310	1986	20,246
1920	689	1966	12,016	1988	21,338
1925	1,069	1968	14,104	1990	22,100
1930	1,412	1970	16,690	1992	22,028
1935	1,655	1972	18,584	1994	22,234
1940	2,150	1974	20,389	1996	23,380
				1998	24,441

The logarithm transformation

The shape of the growth curve for oil production in Figure 2.38 does look like the exponential curve in Figure 2.37, but our eyes are not very good at comparing curves of roughly similar shape. We need a better way to check whether growth is exponential. Our eyes are quite good at judging whether or not points lie along a straight line. So we will apply a mathematical transformation that changes exponential growth into linear growth—and patterns of growth that are not exponential into something other than linear.

The necessary transformation is carried out by taking the *logarithm* of the data points. Use a calculator with a `log` button to compute logarithms. Better yet, most statistical software will calculate the logarithms of all the values of a variable in response to a single command. The essential property of the logarithm for our purposes is that it straightens an exponential growth curve. **If a variable grows exponentially, its logarithm grows linearly.**

EXAMPLE 2.41 Figure 2.39(a) plots the logarithms of the bacteria counts in Figure 2.37. Sure enough, exact exponential growth turns into an exact straight line when we plot the logarithms. After 15 hours, for example, the population contains 2^{15} (32,768) bacteria. The logarithm of 32,768 is 4.515, and this point appears above the 15-hour mark in Figure 2.39(a).

Next, take the logarithms of world oil production from 1880 to 1998 and

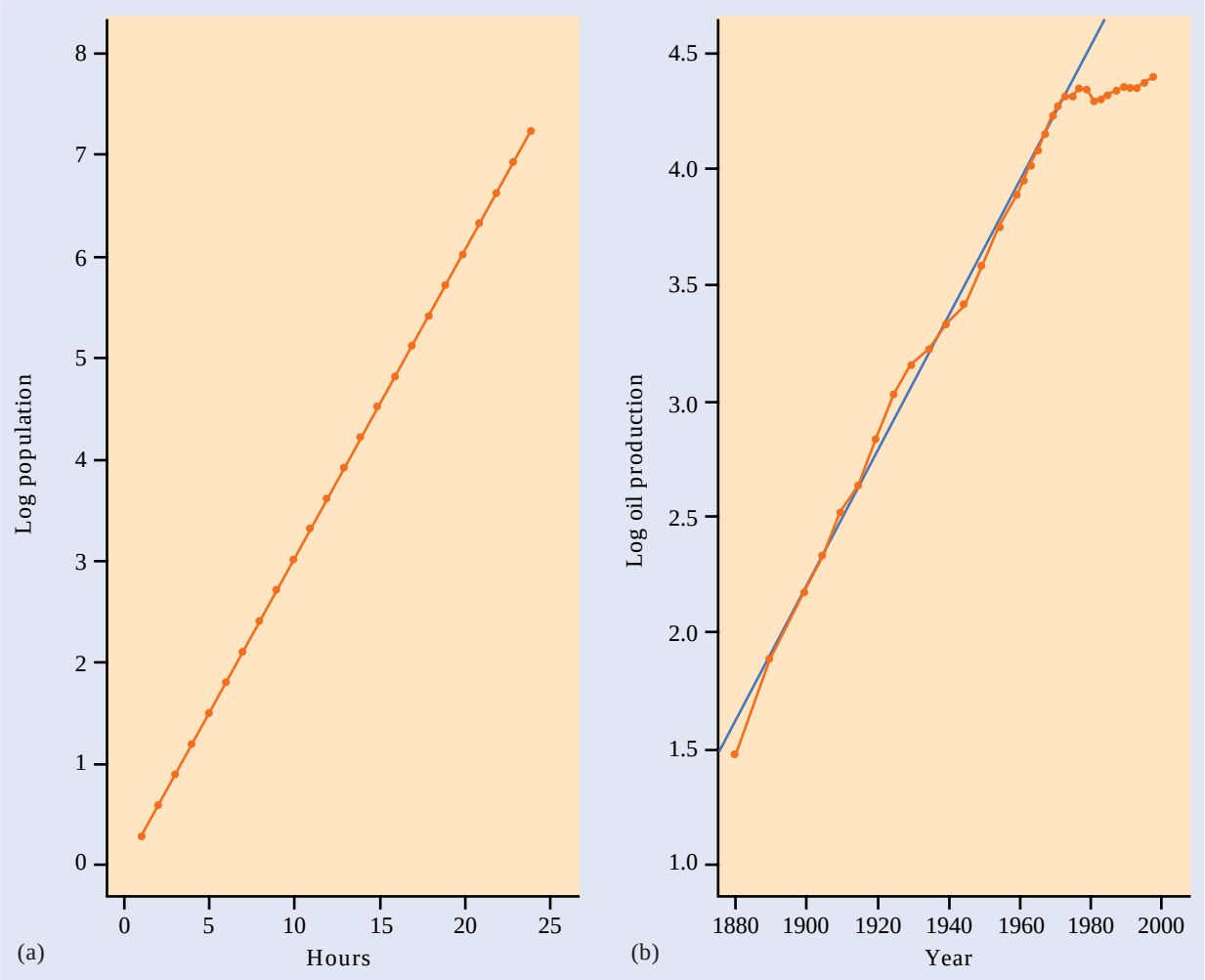


FIGURE 2.39 Taking logarithms turns exponential growth into linear growth. (a) Exact exponential growth: bacteria doubling each hour. (b) Oil production from 1880 to 1998. Growth was approximately exponential until 1973.

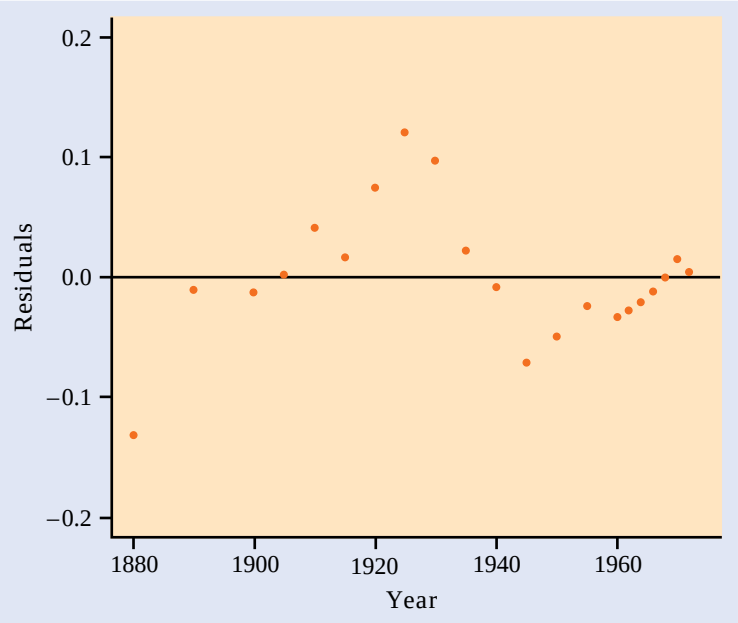


FIGURE 2.40 Residuals from the least-squares regression line of the logarithms of oil production against year, ending in 1972. The residual plot magnifies deviations from the least-squares line and helps us see details of the history of oil production.

plot them against time. Figure 2.39(b) is the result. It looks straight until 1974! The regression line based on the years from 1880 to 1972, added to the plot, describes the long-term trend. The increase in oil production was indeed close to exponential over most of the twentieth century. Perhaps this should have been a warning: in the real world, nothing can grow exponentially for a long period.

Figure 2.40 plots the residuals of the actual logarithms for the years between 1880 and 1972 from the regression line. The residual plot gives us a detailed picture of oil production in these years. Oil production increased more rapidly than the long-term rate in the years between 1900 and the beginning of the Depression in 1929. Production grew more slowly not only during the Depression but also during World War II. That is, the turning points in the residual plot occur in 1925 (the last point before 1929) and 1945 (the end of World War II). Only after the war did oil production return to an above-average growth rate, which lasted until 1973. The production of oil was of course increasing during the entire period up to 1973—the runs of declining residuals show periods of slower growth, not an actual drop in oil production.

Prediction in the exponential growth model

Regression is often used for prediction. When we fit a least-squares regression line, we find the predicted response $\hat{y}$ for any value of the explanatory variable x by substituting our x-value into the equation of the line. In the case of exponential growth,

the logarithms rather than the actual responses follow a linear pattern. To do prediction, we "undo" the logarithm transformation to return to the original units of measurement. The same idea works for any monotonic transformation. There is always exactly one original value behind any transformed value, so we can always go back to our original scale.

EXAMPLE 2.42 Figure 2.41 shows the idea for the logarithm transformation. The logarithm of $x = 2$ is $\log 2 = 0.3010$. The graph shows that we can go "up and over" from $x = 2$ to $\log x = 0.3010$. Because the logarithm function is monotonic, $x = 2$ is the *only* number whose logarithm is 0.3010. So if we are told that $\log x = 0.3010$, we can go "over and down" to recover $x = 2$.

In the form of equations, any number x can be obtained from its common (base 10) logarithm $\log x$ by

$$x = 10^{\log x}$$

If we are told that $\log x = 0.3010$, then

$$x = 10^{\log x} = 10^{0.3010} = 1.9999$$

That's $x = 2$ up to the error caused by rounding $\log x$ to four places.

The regression line for predicting the logarithm of oil production from year x is

$$\log \hat{y} = -52.7 + 0.0289x$$

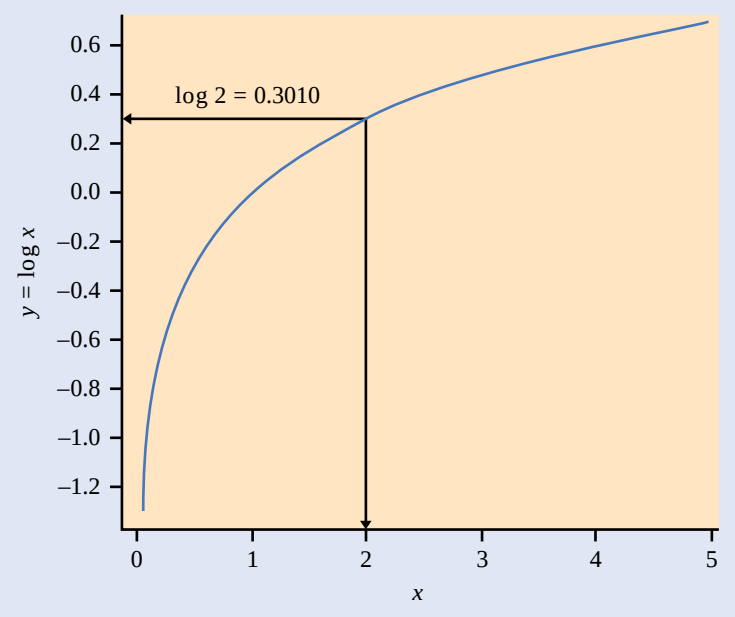


FIGURE 2.41 Any monotonic function has an inverse function that undoes the transformation and returns to the original data. This graph illustrates the process of inverting the logarithm transformation.

This line was fitted to data up to 1972. For 1977, we would predict that

$$\begin{aligned} \log \hat{y} &= -52.7 + (0.0289 \times 1977) \\ &= 4.4353 \end{aligned}$$

The predicted production in the original units (millions of barrels of oil) is therefore

$$\hat{y} = 10^{4.4353} = 27,246$$

The actual production in 1977 was 21,787 million barrels. The prediction is much too high because the exponential growth model fails to fit post-1973 oil production.

Power law models

When you visit a pizza parlor, you order a pizza by its diameter, say 10 inches, 12 inches, or 14 inches. But the amount you get to eat depends on the *area* of the pizza. The area of a circle is π times the square of its radius. So the area of a round pizza with diameter x is

$$\text{area} = \pi r^2 = \pi(x/2)^2 = \pi(x^2/4) = (\pi/4)x^2$$

This is a **power law model** of the form

power law model

$$y = a \times x^p$$

When we are dealing with things of the same general form, whether circles or fish or people, we expect area to go up with the square of a dimension such as diameter or

height. Volume should go up with the cube of a linear dimension. That is, geometry tells us to expect power laws in some settings.

Biologists have found that many characteristics of living things are described quite closely by power laws. There are more mice than elephants, and more flies than mice—the abundance of species follows a power law with body weight as the explanatory variable. So do pulse rate, length of life, the number of eggs a bird lays, and so on. Sometimes the powers can be predicted from geometry, but sometimes they are mysterious. Why, for example, does the rate at which animals use energy go up as the 3/4 power of their body weight? Biologists call this relationship *Kleiber's law.* It has been found to work all the way from bacteria to whales. The search goes on for some physical or geometrical explanation for why life follows power laws. There is as yet no general explanation, but power laws are a good place to start in simplifying relationships for living things.

Exponential growth models become linear when we apply the logarithm transformation to the response variable y. **Power law models become linear when we apply the logarithm transformation to both variables.** Here are the details:

1. The power law model is

$$y = a \times x^p$$

2. Take the logarithm of both sides of this equation. You see that

$$\log y = \log a + p \log x$$

That is, taking the logarithm of both variables straightens the scatterplot of y against x.

3. Look carefully: The *power* p in the power law becomes the *slope* of the straight line that links $\log y$ to $\log x$.

Prediction in power law models

If taking the logarithms of both variables makes a scatterplot linear, a power law is a reasonable model for the original data. We can even roughly estimate what power p the law involves by regressing $\log y$ on $\log x$ and using the slope of the regression line as an estimate of the power. Remember that the slope is only an estimate of the p in an underlying power model. The greater the scatter of the points in the scatterplot about the fitted line, the smaller our confidence that this estimate is accurate.

EXAMPLE 2.43 The magical success of the logarithm transformation in Example 2.37 would not surprise a biologist. We suspect that a power law governs this relationship. Least-squares regression for the scatterplot in Figure 2.33 gives the line

$$\log \hat{y} = 1.01 + 0.72 \times \log x$$

for predicting the logarithm of brain weight from the logarithm of body weight. To undo the logarithm transformation, remember that for com-

mon logarithms with base 10, $y = 10^{\log y}$. We see that

$$\begin{aligned} \hat{y} &= 10^{1.01+0.72 \log x} \\ &= 10^{1.01} \times 10^{0.72 \log x} \\ &= 10.2 \times (10^{\log x})^{0.72} \end{aligned}$$

Because $10^{\log x} = x$, the estimated power model connecting predicted brain weight $\hat{y}$ with body weight x for mammals is

$$\hat{y} = 10.2 \times x^{0.72}$$

Based on footprints and some other sketchy evidence, some people think that a large apelike animal, called Sasquatch or Bigfoot, lives in the Pacific Northwest. His weight is estimated to be about 280 pounds, or 127 kilograms. How big is Bigfoot's brain? Based on the power law estimated from data on other mammals, we predict

$$\begin{aligned} \hat{y} &= 10.2 \times 127^{0.72} \\ &= 10.2 \times 32.7 \\ &= 333.7 \text{ grams} \end{aligned}$$

For comparison, gorillas have an average body weight of about 140 kilograms and an average brain weight of about 406 grams. Of course, Bigfoot may have a larger brain than his weight predicts—after all, he has avoided being captured, shot, or videotaped for many years.

Summary

Nonlinear relationships between two quantitative variables can sometimes be changed into linear relationships by **transforming** one or both of the variables.

The most common transformations belong to the family of **power transformations** t^p. The logarithm $\log t$ fits into the power family at position $p = 0$.

When the variable being transformed takes only positive values, the power transformations are all **monotonic.** This implies that there is an inverse transformation that returns to the original data from the transformed values. The effect of the power transformations on data becomes stronger as we move away from linear transformations ($p = 1$) in either direction.

Transformation is particularly effective when there is reason to think that the data are governed by some mathematical model. The **exponential growth model** $y = ab^x$ becomes linear when we plot $\log y$ against x. The **power law model** $y = ax^p$ becomes linear when we plot $\log y$ against $\log x$.

We can fit exponential growth and power models to data by finding the least-squares regression line for the transformed data, then doing the inverse transformation.

EXERCISES

2.167 Which of these transformations are monotonic increasing? Monotonic decreas-

ing? Not monotonic? Give an equation for each transformation.

(a) You transform height in inches to height in centimeters.

(b) You transform typing speed in words per minute into seconds needed to type a word.

(c) You transform the diameter of a coin to its circumference.

(d) A composer insists that her new piece of music should take exactly 5 minutes to play. You time several performances, then transform the time in minutes into squared error, the square of the difference between 5 minutes and the actual time.

2.168 Suppose that t is an angle, measured in degrees between 0° and 180°. On what part of this range is the function $\sin t$ monotonic increasing? Monotonic decreasing?

2.169 Gordon Moore, one of the founders of Intel Corporation, predicted in 1965 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. Here are data on the dates and number of transistors for Intel microprocessors:[5]

Processor	Date	Transistors	Processor	Date	Transistors
4004	1971	2,250	486 DX	1989	1,180,000
8008	1972	2,500	Pentium	1993	3,100,000
8080	1974	5,000	Pentium II	1997	7,500,000
8086	1978	29,000	Pentium III	1999	24,000,000
286	1982	120,000	Pentium 4	2000	42,000,000
386	1985	275,000			

(a) Explain why Moore's law says that the number of transistors grows exponentially over time.

(b) Make a plot suitable to check for exponential growth. Does it appear that the number of transistors on a chip has in fact grown approximately exponentially?

2.170 A handbook on safe handling of radioactive isotopes gives the following information on the decay of the isotope iodine 125, commonly used in medical and biological experiments:

Days	Activity	Days	Activity
0	1.000	120	0.250
20	0.794	140	0.198
40	0.630	160	0.157
60	0.500	180	0.125
80	0.397	200	0.099
100	0.315	220	0.079

(a) Make plots to decide whether a linear model, an exponential growth or decay model, or a power model best fits these data.

(b) Use regression on transformed data to fit your chosen model. What equation would you use to predict activity after x days?

(c) Do you think that the handbook reports actual laboratory data or data calculated from a theoretical model? Why?

2.171 Return to Moore's law, described in Exercise 2.120.

(a) Find the least-squares regression line for predicting the logarithm of the number of transistors on a chip from the date. Before calculating your line, subtract 1971 from all the dates so that 1971 becomes year 0, 1972 is year 1, and so on.

(b) Suppose that Moore's law is exactly correct. That is, the number of transistors is 2,250 in year 0 (1971) and doubles every 18 months (1.5 years) thereafter. Write the model for predicting transistors in year x after 1971. What is the equation of the line that, according to your model, connects the logarithm of transistors with x? Explain why a comparison of this line with your regression line from (a) shows that, while transistor counts have grown exponentially, they have grown a bit more slowly than Moore's law predicts.

2.172 Here are data for 12 perch caught in a lake in Finland:[6]

Weight (grams)	Length (cm)	Width (cm)	Weight (grams)	Length (cm)	Width (cm)
5.9	8.8	1.4	300.0	28.7	5.1
100.0	19.2	3.3	300.0	30.1	4.6
110.0	22.5	3.6	685.0	39.0	6.9
120.0	23.5	3.5	650.0	41.4	6.0
150.0	24.0	3.6	820.0	42.5	6.6
145.0	25.5	3.8	1000.0	46.6	7.6

(a) Make a scatterplot of weight against length. Describe the pattern you see.

(b) How do you expect the weight of animals of the same species to change as their length increases? Make a transformation of weight that should straighten the plot if your expectation is correct. Plot the transformed weights against length. Is the plot now roughly linear?

2.173 Plot the widths of the 12 perch in the previous problem against their lengths. What is the pattern of the plot? Explain why we should expect this pattern.

2.174 (Exact exponential growth) The common intestinal bacterium *E. coli* is one of the fastest-growing bacteria. Under ideal conditions, the number of *E. coli* in a colony doubles about every 15 minutes until restrained by lack of resources. Starting from a single bacterium, how many *E. coli* will there be in 1 hour? In 5 hours?

2.175 (Exact exponential growth) A paper in a scholarly journal once claimed (I

am not making this up), "Every year since 1950, the number of American children gunned down has doubled."[7] To see that this is silly, suppose that in 1950 just 1 child was "gunned down" and suppose that the paper's claim is exactly right.

(a) Make a table of the number of children killed in each of the next 10 years, 1951 to 1960.

(b) Plot the number of deaths against the year and connect the points with a smooth curve. This is an exponential curve.

(c) The paper appeared in 1995, 45 years after 1950. How many children were killed in 1995, according to the paper?

(d) Take the logarithm of each of your counts from (a). Plot these logarithms against the year. You should get a straight line.

(e) From your graph in (d) find the approximate values of the slope b and the intercept a for the line. Use the equation $y = a + bx$ to predict the logarithm of the count for the 45th year. Check your result by taking the logarithm of the count you found in (c).

2.176 (Exact exponential growth) Maria is given a savings bond at birth. The bond is initially worth \$500 and earns interest at 7.5% each year. This means that the value is multiplied by 1.075 each year.

(a) Find the value of the bond at the end of 1 year, 2 years, and so on up to 10 years.

(b) Plot the value y against years x. Connect the points with a smooth curve. This is an exponential curve.

(c) Take the logarithm of each of the values y that you found in (a). Plot the logarithm $\log y$ against years x. You should obtain a straight line.

2.177 (Exact exponential growth) Fred and Alice were born the same year, and each began life with \$500. Fred added \$100 each year, but earned no interest. Alice added nothing, but earned interest at 7.5% annually. After 25 years, Fred and Alice are getting married. Who has more money?

2.178 The following table gives the resident population of the United States from 1790 to 2000, in millions of persons:

Date	Pop.	Date	Pop.	Date	Pop.	Date	Pop.
1790	3.9	1850	23.2	1910	92.0	1970	203.3
1800	5.3	1860	31.4	1920	105.7	1980	226.5
1810	7.2	1870	39.8	1930	122.8	1990	248.7
1820	9.6	1880	50.2	1940	131.7	2000	281.4
1830	12.9	1890	62.9	1950	151.3		
1840	17.1	1900	76.0	1960	179.3		

(a) Plot population against time. The growth of the American population appears roughly exponential.

(b) Plot the logarithms of population against time. The pattern of growth is now clear. An expert says that "the population of the United States increased exponentially from 1790 to about 1880. After 1880 growth was still approximately exponential, but at a slower rate." Explain how this description is obtained from the graph.

2.179 Do mold colonies grow exponentially? In an investigation of the growth of molds, biologists inoculated flasks containing a growth medium with equal amounts of spores of the mold *Aspergillus nidulans.* They measured the size of a colony by analyzing how much remains of a radioactive tracer substance that is consumed by the mold as it grows. Each size measurement requires destroying that colony, so that the data below refer to 30 separate colonies. To smooth the pattern, we take the mean size of the three colonies measured at each time.[8]

Hours	Colony sizes			Mean
0	1.25	1.60	0.85	1.23
3	1.18	1.05	1.32	1.18
6	0.80	1.01	1.02	0.94
9	1.28	1.46	2.37	1.70
12	2.12	2.09	2.17	2.13
15	4.18	3.94	3.85	3.99
18	9.95	7.42	9.68	9.02
21	16.36	13.66	12.78	14.27
24	25.01	36.82	39.83	33.89
36	138.34	116.84	111.60	122.26

(a) Graph the mean colony size against time. Then graph the logarithm of the mean colony size against time.

(b) On the basis of data like these, microbiologists divide the growth of mold colonies into three phases that follow each other in time. Exponential growth occurs during

only one of these phases. Briefly describe the three phases, making specific reference to the graphs to support your description.

(c) The exponential growth phase for these data lasts from about 6 hours to about 24 hours. Find the least-squares regression line of the logarithms of mean size on hours for only the data between 6 and 24 hours. Use this line to predict the size of a colony 10 hours after inoculation. (The line predicts the logarithm. You must obtain the size from its logarithm.)

2.180 It is easy to measure the "diameter at breast height" of a tree. It's hard to measure the total "aboveground biomass" of a tree, because to do this you must cut and weigh the tree. The biomass is important for studies of ecology, so ecologists commonly estimate it using a power law. Combining data on 378 trees in tropical rain forests gives this relationship between biomass y measured in kilograms and diameter x measured in centimeters:[9]

$$\log_e y = -2.00 + 2.42 \log_e x$$

Note that the investigators chose to use *natural logarithms,* with base $e = 2.71828$, rather than common logarithms with base 10.

(a) Translate the line given into a power model. Use the fact that for natural logarithms,

$$y = e^{\log_e y}$$

(b) Estimate the biomass of a tropical tree 30 centimeters in diameter.

TABLE 2.11

Body weight and lifetime for several species of mammals

Species	Weight (kg)	Life span (years)
Baboon	32	20
Beaver	25	5
Cat, domestic	2.5	12
Chimpanzee	45	20
Dog	8.5	12
Elephant	2800	35
Goat, domestic	30	8
Gorilla	140	20
Grizzly bear	250	25
Guinea pig	1	4
Hippopotamus	1400	41
Horse	480	20
Lion	180	15
Mouse, house	0.024	3
Pig, domestic	190	10
Red fox	6	7
Sheep, domestic	30	12

2.181 Find the correlation between the logarithm of mean size and hours for the data between 6 and 24 hours in Exercise 2.130. Make a scatterplot of the logarithms of the individual size measurements against hours for this same period and find the correlation. Why do we expect the second r to be smaller? Is it in fact smaller?

2.182 Table 2.11 gives the average weight and average life span in captivity for several species of mammals. Some writers on power laws in biology claim that life span depends on body weight according to a power law with power $p = 0.2$. Fit a power law model to these data (using logarithms). Does this small set of data appear to follow a power law with power close to 0.2? Use your fitted model to predict the average life span for humans (average weight 65 kilograms). Humans are an exception to the rule.

2.183 Bigger people are generally stronger than smaller people, though there's a lot of individual variation. Let's find a theoretical model. Body weight increases as the cube of height. The strength of a muscle increases with its cross-sectional area, which we expect to go up as the square of height. Put these together: What power law should describe how muscle strength increases with weight?

2.184 Let's apply your result from the previous problem. Graph the power law relation between strength and body weight for weights from (say) 1 to 1000. (Constants in the power law just reflect the units of measurement used, so we can ignore them.) Use the graph to explain why a person 1 million times as heavy as an ant can't lift a

million times as much as an ant can lift.

2.185 Physiologists say that the resting heart rate of humans is related to our body weight by a power law. Specifically, average heart rate y (beats per minute) is found from body weight x (kilograms) by[10]

$$y = 241 \times x^{-1/4}$$

Let's try to make sense of this. Kleiber's law says that energy use in animals, including humans, increases as the 3/4 power of body weight. But the weight of human hearts and lungs and the volume of blood in the body are directly proportional to body weight. Given these facts, you should not be surprised that heart rate is proportional to the $-1/4$ power of body weight. Why not?

2.186 Galileo studied motion by rolling balls down ramps. Newton later showed how Galileo's data fit his general laws of motion. Imagine that you are Galileo, without Newton's laws to guide you. He rolled a ball down a ramp at different heights above the floor and measured the horizontal distance the ball traveled before it hit the floor. Here are Galileo's data when he placed a horizontal shelf at the end of the ramp so that the ball is moving horizontally when it starts to fall. (We won't try to describe the obscure seventeenth-century units Galileo used to measure distance.)[11]

Distance	Height
1500	1000
1340	828
1328	800
1172	600
800	300

Plot distance y against height x. The pattern is very regular, as befits data described by a physical law. We want to find distance as a function of height. That is, we want to transform x to straighten the graph.

(a) Think before you calculate: Will powers x^p for $p < 1$ or $p > 1$ tend to straighten the graph. Why?

(b) Move along the ladder of transformations in the direction you have chosen until the graph is nearly straight. What transformation do you suggest?

2.187 Table 2.12 gives data on the mean number of seeds produced in a year by several common tree species and the mean weight (in milligrams) of the seeds produced. (Some species appear twice because their seeds were counted in two locations.) We might expect that trees with heavy seeds produce fewer of them, but what is the form of the relationship?[12]

(a) Make a scatterplot showing how the weight of tree seeds helps explain how many seeds the tree produces. Describe the form, direction, and strength of the relationship.

(b) If a power law holds for this relationship, the logarithms of the original data will

display a linear pattern. Use your calculator or software to obtain the logarithms of both the seed weights and the seed counts in Table 2.12. Make a new scatterplot using these new variables. Now what are the form, direction, and strength of the relationship?

2.188 Long ago, Thomas Malthus (1766–1834) claimed that population grows exponentially while food supplies grow only linearly. Result: starvation holds down population. Table 2.13 gives data on the population of the world, as estimated by the U.S. Census Bureau. Do a careful analysis, using our study of world petroleum production as a model, and write a brief discussion of the pattern of increase in the world's population since 1950. Has population grown exponentially since 1950? Has population growth slowed in recent years?

2.189 Expose marine bacteria to X-rays for time periods from 1 to 15 minutes. Here are the number of surviving bacteria (in hundreds) on a culture plate after each exposure time:[13]

Time t	Count y	Time t	Count y
1	355	9	56
2	211	10	38
3	197	11	36
4	166	12	32
5	142	13	21
6	106	14	19
7	104	15	15
8	60		

Theory suggests an exponential growth or decay model. Do the data appear to conform to this theory?

2.190 Electronic fund transfers, from bank automatic teller machines and the use of debit cards by consumers, have grown rapidly in the United States. Here are data on the number of such transfers (in millions):[14]

Year	EFT	Year	EFT	Year	EFT
1985	3,579	1991	6,642	1996	11,780
1987	4,108	1992	7,537	1997	12,580
1988	4,581	1993	8,135	1998	13,160
1989	5,274	1994	9,078	1999	13,316
1990	5,942	1995	10,464		

Write a clear account of the pattern of growth of electronic transfers over time, sup-

porting your description with plots and calculations as needed. Has the pattern changed in the most recent years?

TABLE 2.12

Count and weight of seeds produced by common tree species

Tree species	Seed count	Seed weight (mg)	Tree species	Seed count	Seed weight (mg)
Paper birch	27,239	0.6	American beech	463	247
Yellow birch	12,158	1.6	American beech	1,892	247
White spruce	7,202	2.0	Black oak	93	1,851
Engelmann spruce	3,671	3.3	Scarlet oak	525	1,930
Red spruce	5,051	3.4	Red oak	411	2,475
Tulip tree	13,509	9.1	Red oak	253	2,475
Ponderosa pine	2,667	37.7	Pignut hickory	40	3,423
White fir	5,196	40.0	White oak	184	3,669
Sugar maple	1,751	48.0	Chestnut oak	107	4,535
Sugar pine	1,159	216.0			

TABLE 2.13

Midyear world population, billions

Year	Population	Year	Population	Year	Population
1950	2.555	1968	3.558	1986	4.938
1952	2.635	1970	3.708	1988	5.110
1954	2.728	1972	3.862	1990	5.284
1956	2.833	1974	4.015	1992	5.450
1958	2.945	1976	4.160	1994	5.611
1960	3.039	1978	4.305	1996	5.769
1962	3.136	1980	4.457	1998	5.925
1964	3.277	1982	4.613	2000	6.080
1966	3.416	1984	4.774	2002	6.234

Supplement to Section 4.5

Decision Analysis

Perhaps the most straightforward view of statistical inference regards the goal of inference as *making decisions in the presence of uncertainty.* Since uncertainty can be described by probability, the use of probability in inference is essential. We will illustrate the nature of inference as decision by several examples.

Decisions based on probabilities

One possible goal of decision-making in the presence of uncertainty is to make the probability of a favorable outcome as large as possible. Here is an example of such a problem.

EXAMPLE 4.49 The kidneys of a patient with end-stage kidney disease will not support life. If the patient is to survive, the available choices are a kidney transplant or regular hemodialysis (use of a kidney machine several times a week). Lynn is faced with this choice. Both treatments are risky. Her doctor gives her the following information for patients in her condition: About 68% of dialysis patients survive for 5 years. Of transplant patients, about 48% survive with the transplanted kidney for five years, 43% must undergo regular dialysis because the transplanted kidney fails, and the remaining 9% do not survive the transplant. Of

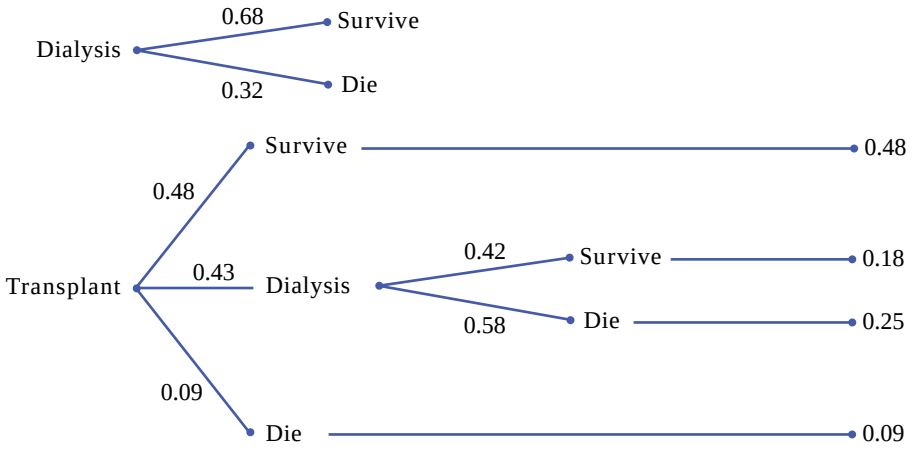


FIGURE 4.20 Tree diagram for the kidney failure decision problem in Example 4.47.

those transplant patients who return to dialysis, about 42% survive 5 years. Lynn wishes to make her chances of living for 5 years as great as possible. Which treatment should she choose?[15]

We can organize the information provided by the doctor in the *tree diagram* of the possible outcomes in Figure 4.20. Each path through the tree represents a possible outcome of Lynn's case. The probability written beside each branch after the first stage is the *conditional* probability of the next step given that Lynn has reached this point. For example, 0.68 is the probability that a dialysis patient survives 5 years. The conditional probability that a patient survives 5 years on dialysis given that the patient first had a transplant then returned to dialysis because the transplant failed is different. It is 0.42, and appears on a different branch of the tree. Study the tree to convince yourself that it organizes all the information available.

The multiplication rule for probabilities says that *the probability of reaching the end of a branch in the tree is the product of the probabilities of all the steps along that branch.* These probabilities are written at the end of each branch in Figure 4.20. For example, the event that a transplant patient returns to dialysis and then survives 5 years is

$$A = \{D \text{ and } S\}$$

where

D={transplant patient returns to dialysis}

S={patient survives 5 years}

Therefore

$$\begin{aligned} P(A) &= P(D \text{ and } S) \\ &= P(S|D)P(D) \\ &= (0.42)(0.43) = 0.18 \end{aligned}$$

Lynn can now compare her probability of surviving 5 years with regular dialysis with the probability of surviving 5 years with a transplant. For dialysis, the survival probability is 0.68. A transplant patient can either survive with the transplant functioning, or survive a failed transplant by returning to dialysis. These are *disjoint* events, represented by separate branches of the tree. They have probabilities 0.48 and 0.18, respectively. The addition rule for probability says that the overall probability of surviving 5 years is the sum, $0.48 + 0.18 = 0.66$. This is slightly less than the survival probability under dialysis, so dialysis is the better choice.

The medical decision analysis for Example 4.49 illustrates a decision problem in which the goal is to maximize the probability of a desired outcome. The decision tree organizes the different paths that can lead to the final outcome, and makes the calculation of outcome probabilities easy once all the necessary conditional probabilities are in place. We can then compare the probability of the desired outcome under each of the possible decisions. In Example 4.49, there are only two decisions to be compared, dialysis or transplant. Other decision problems may have more alternatives.

Where do the conditional probabilities in Example 4.49 come from? They are based in part on data, that is, on studies of many kidney disease patients. But an

individual's chances of survival depend on her age, general health, and other factors. Lynn's doctor considered her individual situation before giving these particular probabilities. It is characteristic of most decision analysis problems that *personal probabilities* are used to describe the uncertainty of an informed decision-maker.

Decisions based on mean utility

On reconsidering, Lynn realizes that the mere probability of survival is not an adequate basis for her decision. The better quality of life afforded by a successful transplant should be taken into account. Other factors such as cost are also important. Suppose that Lynn could represent the overall benefit of each possible outcome by a number. Surviving 5 years on dialysis might get a 50, and surviving 5 years with a working transplant an 80. These numbers are called *utilities.* Since the outcome is uncertain, Lynn should make the decision that makes the *mean* of the utility as large as possible. This is a second—and more important—type of decision analysis.

It is hard to assign a numerical utility that measures the benefits of a medical outcome. This is true because there are many factors to consider, and most do not have obvious numerical values. But in some other situations, the utilities of outcomes are measured in dollars, a numerical value that everyone understands. Decision analysis based on maximizing mean dollar returns is widely used in business planning.

EXAMPLE 4.50 Zipdrive, Inc. has developed a new disk drive for small computers. The demand for the new product is uncertain, but can

be described as "high" or "low" in any one year. The product is expected to be obsolete after four years. Management must decide whether to build a plant or to contract with a factory in Hong Kong to manufacture the new drive. Building a plant will be very profitable if demand remains high, but could cause a loss if demand drops in future years.

After careful study of the market and of all relevant costs, Zipdrive's planning office provides the following information. Three situations are possible,

S_1 = First year demand high, next three years demand high

S_2 = First year demand high, next three years demand low

S_3 = First year demand low, next three years demand low

The best estimate of the probabilities of these cases is

Situation	S_1	S_2	S_3
Probability	0.5	0.4	0.1

Finally, the total profits for the four-year period depend on the decision made and the demand that is experienced. The profits (in millions of dollars) in each case can be forecast to be as follows

Decision	S_1	S_2	S_3
Build	56	10	-20
Contract	30	20	10

Which decision is more promising?

We can easily make a tree diagram for this decision problem, but because the outcome involves only a single stage (one of the three demand situations), a tree diagram is not as helpful as in Example 4.49. Let's just compute the mean profit for both decisions. If a new plant is built, the mean profit is

$$\mu = (56)(0.5) + (10)(0.4) + (-20)(0.1) = \$30 \text{ million}$$

Contracting the work to Hong Kong gives a mean profit of

$$\mu = (30)(0.5) + (20)(0.4) + (10)(0.1) = \$24 \text{ million}$$

Building a new plant offers the higher profit, averaged over the uncertain future demand.

The probabilities for the different levels of demand are again personal probabilities that reflect the judgment of experienced managers. Often the manager who developed the new product will be more optimistic than the sales manager. In that case, Zipdrive can calculate the mean profits twice, using both managers' probabilities. This is called a sensitivity analysis, because it shows whether the recommended decision changes when we change the probabilities. In practice (where decision problems are much more complex than our examples) the major value of decision analysis is that it clarifies the conditions under which each choice is best.

As it turns out, the president of Zipdrive is worried about the possibility of losing \$20 million if demand is low, so he decides to contract the manufacturing to Hong

Kong even though that choice has lower mean profits. The president's utility is not the same as profit in dollars, because he is reluctant to risk a loss. Even in a business decision problem, measuring utilities is difficult.

EXERCISES

4.150 John has coronary artery disease. He and his doctor must decide between medical management of the disease and heart bypass surgery. John is concerned about quality of life as well as length of life. He wants to maximize the probability of surviving at least 5 years and being able to carry on moderate activity during that time. Call this event A. For patients of John's age and condition, the doctor estimates that under medical management, $P(A) = 0.7$. If John has bypass surgery,

- There is probability 0.05 that he will not survive, probability 0.10 that he will survive with major complications, and probability 0.85 that he will survive without serious complications.
- If he survives with complications, the conditional probability of A is 0.73. If he survives without complications, the conditional probability of A is 0.76.

Find $P(A)$ if John has surgery and compare it with $P(A) = 0.7$ for medical management. What should John do?[16]

4.151 A snowmobile manufacturer wonders whether the coming winter will be severe, normal, or mild. Weather records show that the probability of a normal winter is 0.5

and that mild and severe winters each occur with probability 0.25. The manufacturer knows what production schedule would be best for each type of winter, and can predict the profit of each schedule in all three types of winter. Of course, the profit will be smaller if the production schedule doesn't match the winter. Here are the predicted profits, in units of $100,000:

	Winter		
Decision	Severe	Normal	Mild
Severe schedule	120	50	−10
Normal schedule	100	100	50
Mild schedule	80	80	80

Find the mean profits for each production schedule. Which choice makes the mean profits largest?

4.152 The snowmobile manufacturer in the previous exercise decides not to use the long-term probabilities of the three types of winter. Instead, he relies on a forecast that the probability is 0.75 that the coming winter will be severe, 0.20 that it will be normal, and 0.05 that it will be mild. Which production schedule will maximize mean profits based on these probabilities?

4.153 A repairman receives a fixed fee of 5 to examine a radar receiver that isn't working. He knows that the problem is either in the microwave amplifier or in the power supply. His cost to examine the amplifier is 3 and his cost for examining the power supply is 1. He must decide which to examine first—and if he guesses wrong,

he must then examine the other component.

(a) Fill in the repairman's profits for each situation:

	Problem location	
Decision	Amplifier	Power supply
Amplifier first		
Power supply first		

(b) From long experience, the repairman knows that 80% of problems like this one are due to a defect in the amplifier. Which component should he examine first?

4.154 The owner of a Mandarin restaurant is considering preparing Peking ducks (which require a day's advance work) to satisfy customers who don't order in advance. Based on his experience, here are personal probabilities for the number of requests per day for Peking duck without advance orders:

Requests	0	1	2	3	4
Probability	0.3	0.2	0.2	0.2	0.1

If he sells a duck, his profit is \$10. Each duck prepared but not sold costs him \$5. The restaurant owner can decide to prepare 0, 1, 2, 3, or 4 ducks in advance. Which decision makes the mean profit largest?

4.155 Here is a simplified version of one of the first uses of decision analysis in anthropology.[17] Fishing canoe captains in a village in Jamaica set fish traps that they later draw up and reset. There are two fishing areas, the inner and outer banks. The outer banks yield more and better fish, but 25% of the time there are strong

currents that can destroy traps. The inner banks yield fewer fish, but never have these currents. An anthropologist observed that many captains set one-third of their traps on the inner banks and two-thirds on the outer banks. After careful study, the anthropologist found that average monthly income for a captain (in British pounds) is as follows:

Decision	Current	No current
All traps inside	17.3	11.5
One-third inside	5.2	17.0
All traps outside	−4.4	20.6

Which trap location decision should a fishing captain make?

NOTES

1. Data from G. A. Sacher and E. F. Staffelt, "Relation of gestation time to brain weight for placental mammals: implications for the theory of vertebrate growth," *American Naturalist,* 108 (1974), pp. 593–613. We found these data in F. L. Ramsey and D. W. Schafer, *The Statistical Sleuth: A Course in Methods of Data Analysis,* Duxbury, 1997.

2. There are several mathematical ways to show that $\log t$ fits into the power family at $p = 0$. Here's one. For powers $p \neq 0$, the indefinite integral $\int t^{p-1}dt$ is a multiple of t^p. When $p = 0$, $\int t^{-1}dt$ is $\log t$.

3. Data from the World Bank's *1999 World Development Indicators.* Life expectancy is estimated for 1997 and GDP per capita (purchasing power parity basis) for 1998.

4. Data from the Energy Information Administration, recorded in Robert H. Romer, *Energy: An Introduction to Physics,* W. H. Freeman, 1976, for 1880 to 1972, and from the EIA's *International Energy Annual* for later years.

5. From the Intel Web site, `www.intel.com/research/silicon/mooreslaw.htm`.

6. Data on a sample of 12 of 56 perch in a data set contributed to the *Journal of Statistics Education* data archive (`www.amstat.org/publications/jse`) by Juha Puranen of the University of Helsinki.

7. From Joel Best, *Damned Lies and Statistics: Untangling Numbers from the Media,*

Politicians, and Activists, University of California Press, 2001.

8. Similar experiments are described by A. P. J. Trinci, "A kinetic study of the growth of *Aspergillus nidulans* and other fungi," *Journal of General Microbiology,* 57 (1969), pp. 11–24. These data were provided by Thomas Kuczek, Purdue University.

9. Jérôme Chave, Bernard Riéra, and Marc-A. Dubois, "Estimation of biomass in a neotropical forest of French Guiana: spatial and temporal variability," *Journal of Tropical Ecology,* 17 (2001), pp. 79–96.

10. The power law connecting heart rate with body weight was found online at "The Worldwide Anaesthetist," `www.anaesthetist.com`. Anesthesiologists are interested in power laws because they must judge how drug doses should increase in bigger patients.

11. Data from Stillman Drake, *Galileo at Work,* University of Chicago Press, 1978. We found these data in D. A. Dickey and J. T. Arnold, "Teaching statistics with data of historic significance," *Journal of Statistics Education,* 3 (1995), `www.amstat.org/publications/jse`.

12. Data from many studies compiled in D. F. Greene and E. A. Johnson, "Estimating the mean annual seed production of trees," *Ecology,* 75 (1994), pp. 642–647.

13. S. Chatterjee and B. Price, *Regression Analysis by Example,* Wiley, 1977.

14. From several editions of the *Statistical Abstract of the United States.*

15. The probabilities in this example are based on the article by Benjamin A. Barnes, "An overview of the treatment of end-stage renal disease and a consideration of some of the consequences," in J. P. Bunker, B. A. Barnes, and F. W. Mosteller (eds), *Costs, Risks and Benefits of Surgery,* Oxford University Press, New York, 1977, pp. 325–341. See this article for a more realistic analysis.

16. Based loosely on M. C. Weinstein, J. S. Pliskin, and W. B. Stason, "Coronary artery bypass surgery: decision and policy analysis" in the volume cited in Note 15.

17. William Davenport, "Jamaican fishing: a game theory analysis," in Sidney Mintz (ed.), *Papers in Caribbean Anthropology,* Human Relations Area Files, 1970.

Table 2.10 Annual world crude oil production, 1880–1998

(millions of barrels)

Year	Mbbl.	Year	Mbbl.	Year	Mbbl.
1880	30	1945	2,595	1976	20,188
1890	77	1950	3,803	1978	21,922
1900	149	1955	5,626	1980	21,722
1905	215	1960	7,674	1982	19,411
1910	328	1962	8,882	1984	19,837
1915	432	1964	10,310	1986	20,246
1920	689	1966	12,016	1988	21,338
1925	1,069	1968	14,104	1990	22,100
1930	1,412	1970	16,690	1992	22,028
1935	1,655	1972	18,584	1994	22,234
1940	2,150	1974	20,389	1996	23,380
				1998	24,441

Table 2.11 Body weight and lifetime for several species of mammals

Species	Weight (kg)	Life span (years)
Baboon	32	20
Beaver	25	5
Cat, domestic	2.5	12
Chimpanzee	45	20
Dog	8.5	12
Elephant	2800	35
Goat, domestic	30	8
Gorilla	140	20
Grizzly bear	250	25
Guinea pig	1	4
Hippopotamus	1400	41
Horse	480	20
Lion	180	15
Mouse, house	0.024	3
Pig, domestic	190	10
Red fox	6	7
Sheep, domestic	30	12

Table 2.12 Count and weight of seeds produced by common tree species

Tree species	Seed count	Seed weight (mg)	Tree species	Seed count	Seed weight (mg)
Paper birch	27,239	0.6	American beech	463	247
Yellow birch	12,158	1.6	American beech	1,892	247
White spruce	7,202	2.0	Black oak	93	1,851
Engelmann spruce	3,671	3.3	Scarlet oak	525	1,930
Red spruce	5,051	3.4	Red oak	411	2,475
Tulip tree	13,509	9.1	Red oak	253	2,475
Ponderosa pine	2,667	37.7	Pignut hickory	40	3,423
White fir	5,196	40.0	White oak	184	3,669
Sugar maple	1,751	48.0	Chestnut oak	107	4,535
Sugar pine	1,159	216.0			

Table 2.13 Midyear world population, billions

Year	Population	Year	Population	Year	Population
1950	2.555	1968	3.558	1986	4.938
1952	2.635	1970	3.708	1988	5.110
1954	2.728	1972	3.862	1990	5.284
1956	2.833	1974	4.015	1992	5.450
1958	2.945	1976	4.160	1994	5.611
1960	3.039	1978	4.305	1996	5.769
1962	3.136	1980	4.457	1998	5.925
1964	3.277	1982	4.613	2000	6.080
1966	3.416	1984	4.774	2002	6.234