4.1 Transforming Relationships

http://www.ruf.rice.edu/~lane/stat_sim/transformations/index.html

http://tools.google.com/gapminder/

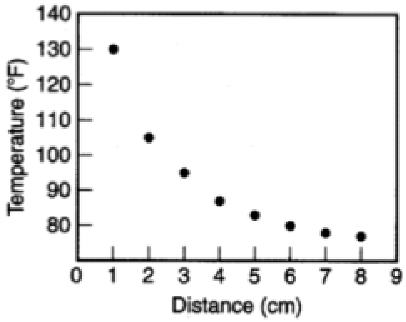
Transform data to make data more

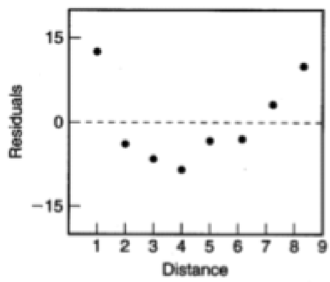
- closely approximate a theoretical distribution
- evenly spread out/constant in variance
- symmetric
- linear

The table shows the temperature (*Temp*) of an instrument measured as its distance (*Dist*) from a heat source is varied. Although calculation would yield $r = -.894$, the scatterplot shows clearly that the data do not have a linear relationship.

Distance (cm)	Temperature (°F)
1	130
2	105
3	95
4	87
5	83
6	80
7	78
8	77

Scatterplot:



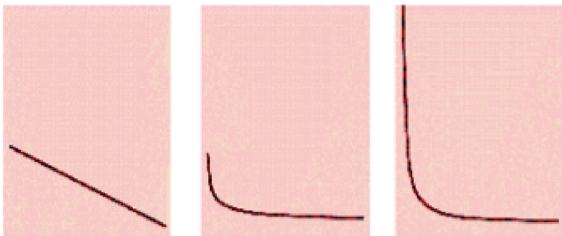


monotonic functions: only increasing or only decreasing

monotonic increasing

Linear, positive slope Square Logarithm

monotonic decreasing



Linear, negative slope Reciprocal square root Reciprocal

Exercise 4.1
Is it monotonic increasing, monotonic decreasing, or not monotonic? Give an equation for each transformation.

a. Transform height from inches to centimeters.

b. Transform words per minute into seconds needed to type a word.

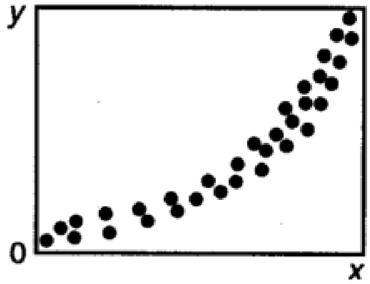
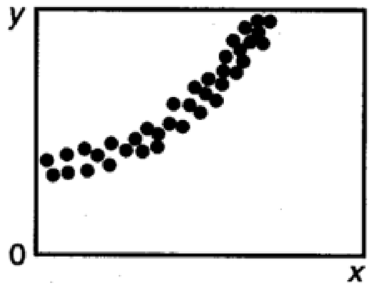
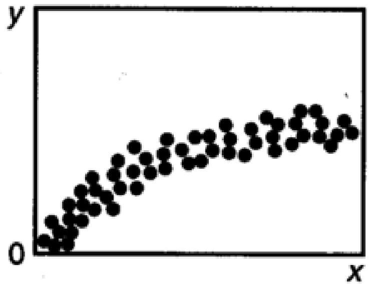
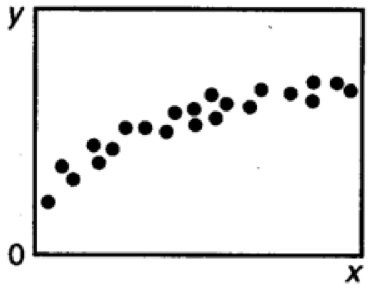
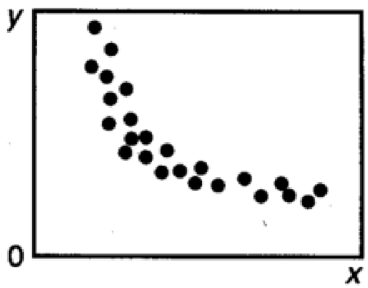
c. Transform diameter to circumference.

d. A piece of music should take exactly 5 minutes to play. Time several performances, then transform the time into squared error, $(5 \text{ minutes - time to play})^2$.

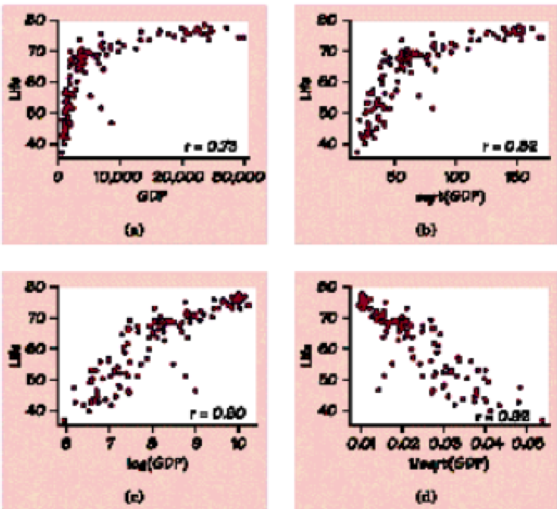
Try some very general suggestions:

clustered near origin and increasing ⇒	log x and log y
concave down and increasing ⇒	$\sqrt{x}$ or log x
concave up with rapid growth ⇒	log y
concave up with rapid growth but clustered at right ⇒	log x and log y
concave up and decreasing ⇒	1/x and 1/y

Any model may be subject to improvement.

Intuitive curve of best fit	Example scatterplot	Suggested transformation
Contains (0, 0) and appears to be a power curve, or a curve asymptotic to both horizontal and vertical axes.		$(x_i, y_i) \rightarrow (\ln x_i, \ln y_i)$ $x_i > 0$ $y_i > 0$
Contains a nonzero *y*-intercept and appears exponential (either growth or decay).		$(x_i, y_i) \rightarrow (x_i, \ln y_i)$ $y_i > 0$
Contains (0, 0) and appears logarithmic.		$(x_i, y_i) \rightarrow (\sqrt{x_i}, y_i)$ $x_i \geq 0$
Contains a nonzero *y*-intercept and appears logarithmic.		$(x_i, y_i) \rightarrow (\ln x_i, y_i)$ $x_i > 0$
Has nonzero horizontal and vertical asymptotes.		$(x_i, y_i) \rightarrow \left(\frac{1}{x_i}, \frac{1}{y_i}\right)$ $x_i \neq 0$ $y_i \neq 0$

original scatterplot



transformation by square root function

transformation by the log function

transformation by the reciprocal square root function

Which model?

Linear:
- each term changes by adding a constant
- look for a (nearly) common difference between consecutive terms

Exponential:
- each term changes by multiplying by a constant
- look for a (nearly) common ratio of consecutive terms

Exercise 4.3

Bigger people are generally stronger than smaller people, though there's a lot of individual variation. Let's find a theoretical model. Body weight (based on volume) increases as the cube of height (length x width x 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?

$w = c_1 h^3 \rightarrow h^3 = \frac{w}{c_1} \rightarrow h = \sqrt[3]{\frac{w}{c_1}}$

$S = c_2 h^2$

$S = c_2 \left(\sqrt[3]{\frac{w}{c_1}}\right)^2 \quad S = c_2 \cdot \frac{w^{2/3}}{c_1^{2/3}}$

$S = C\, W^{2/3}$

Weight = c_1(height)^3 and Strength = c_2(height)^2; therefore, by substitution, we have Strength = c(weight)$^{2/3}$ where c is a constant.

Exercise 4.5

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

define variables x and y

Kleiber's law says that energy use... increases as the 3/4 power of body weight.

energy is proportional to weight

Weight of human hearts and lungs and volume of blood are directly proportional to body weight.

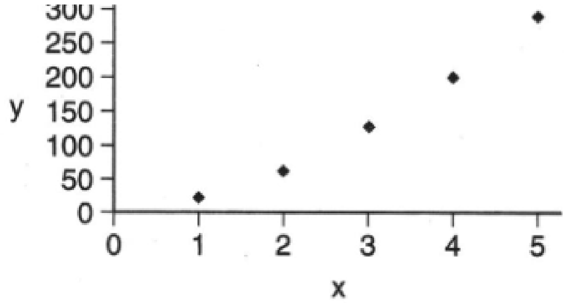
energy is proportional to blood volume and heart rate

blood volume is proportional to weight

Justify the rule $y=241x^{-1/4}$.

put these all together

x	1	2	3	4	5
y	20	60	120	190	280



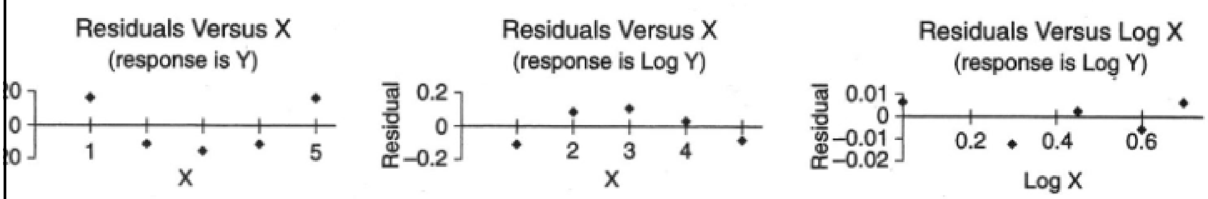
Answer: A linear fit to x and y gives $\hat{y} = 65x - 61$ with $r = .99$. A linear fit to x and log y gives $\log y = 0.279x + 1.139$ with $r = .98$. This results in an *exponential* relationship:

$$\hat{y} = 10^{0.279x+1.139} = 13.77(10^{0.279x}) = 13.77(1.901^x)$$

A linear fit to $\log x$ and $\log y$ gives $\log y = 1.639 \log x + 1.295$, also with $r = .99$. This results in a *power* relationship:

$$\hat{y} = 10^{1.639 \log x + 1.295} = 19.72(x^{1.639})$$

All three models give high correlation and are reasonable fits. Further analysis can be done by examining the residual plots:



The first two residual plots have distinct curved patterns. Among the above three models, the power model, $\hat{y} = 19.72(x^{1.639})$, appears to be best.

Steps in transforming data

1. Plot the data with Excel or the calculator.

2. If linear, find the LSRL, r, r squared, and examine the residual plot to verify a linear model is appropriate.

3. If not linear, use a regression equation from transformed data to find the model in terms of x and y. Transform the data based on whether it is...
...power: Using (log x, log y) may straighten it.
...exponential: Using (x, log y) may straighten it.
...logarithmic: Using (log x, y) may straighten it.
... related to some power: Using the "ladder of power transformations" may straighten it.

Mystery Data

t x	d y
0.2409	0.3871
0.6152	0.7323
1	1
1.881	1.524
11.86	5.203
29.46	9.555
84.01	19.22
164.8	30.11
247.7	39.81

Steps in transforming data

1. Plot the data with Excel or the calculator.

2. If linear, find the LSRL, r, r squared, and examine the residual plot to verify a linear model is appropriate.

3. If not linear, use a regression equation from transformed data to find the model in terms of x and y. Transform the data based on whether it is...
...power: Using (log x, log y) may straighten it.
...exponential: Using (x, log y) may straighten it.
...logarithmic: Using (log x, y) may straighten it.
... related to some power: Using the "ladder of power transformations" may straighten it.

year	stores
1971	1
1987	17
1988	33
1989	55
1990	84
1991	116
1992	165
1993	272
1994	425
1995	676
1996	1015
1997	1412
1998	1886
1999	2135
2000	3501
2001	4709
2002	5886
2003	7225
2004	8337

Steps in transforming data

1. Plot the data with Excel or the calculator.

2. If linear, find the LSRL, r, r squared, and examine the residual plot to verify a linear model is appropriate.

3. If not linear, use a regression equation from transformed data to find the model in terms of x and y. Transform the data based on whether it is...
...power: Using (log x, log y) may straighten it.
...exponential: Using (x, log y) may straighten it.
...logarithmic: Using (log x, y) may straighten it.
... related to some power: Using the "ladder of power transformations" may straighten it.

http://www.starbucks.com/aboutus/timeline.asp

$$\log \hat{y} = -265.06 + .134x$$

$$10^{\log \hat{y}} = 10^{(-265.06 + .134x)}$$

$$y = \underbrace{10^{-265.06}}_{a} \cdot \underbrace{10^{.134x}}_{b}$$

LinRegTTest
y=a+bx
β≠0 and ρ≠0
↑p=1.002792E-13
df=17
a=-265.0699819
↓b=.1342424422

$$y = a b^x$$

$$a^n \cdot a^m = a^{n+m}$$

For some chemical:

rate	concentration
1.1	0.0794328226
4.7	0.0000199498
4.1	0.0000794238
3.6	0.0002511902
4.3	0.0000501201
3	0.0010000400
2	0.0100005010
5.1	0.0000079483
1	0.1000000030
2.3	0.0050119023
1.6	0.0251188653
1.9	0.0125892491

Steps in transforming data

1. Plot the data with Excel or the calculator.

2. If linear, find the LSRL, r, r squared, and examine the residual plot to verify a linear model is appropriate.

3. If not linear, use a regression equation from transform data to find the model in terms of x and y. Transform the data based on whether it is...
...power: Using (log x, log y) may straighten it.
...exponential: Using (x, log y) may straighten it.
...logarithmic: Using (log x, y) may straighten it.
... related to some power: Using the "ladder of power transformations" may straighten it.