A bit of a primer on Nonlinear regression: Polynomials and transformations

by Chris Olsen
COlsen@mchsi.com

Revised 11/30/2019

When points in a scatterplot exhibit a linear pattern and the residual plot does not reveal problems with the linear fit, the least-squares line is an appropriate way to summarize and analyze the relationship between *x* and *y*. A linear relationship is easy to interpret, departures from the line are easily detected, and using the line to predict *y* from our knowledge of *x* is straightforward. Occasionally a scatterplot or residual plot exhibits a curved pattern, indicating a more complicated relationship between *x* and *y*. Nonlinear regression is one of the tools used for addressing these complications. The general ideas of nonlinear fits are the same as you have experienced with straight line fits:

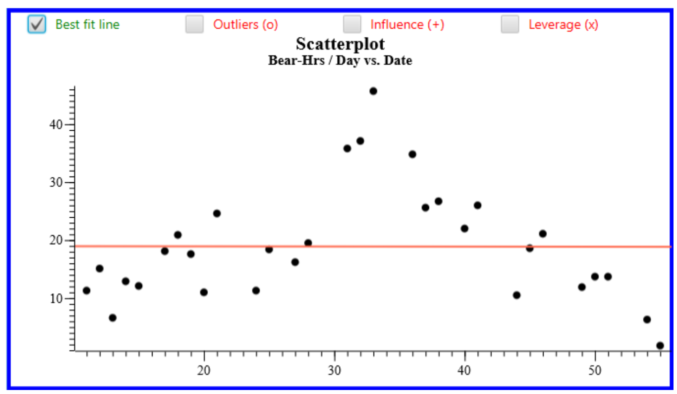
What you have done with straight line fits	What you will do with non-straight line fits
Capture linear relationships using a linear function	Capture nonlinear relationships using a non-linear function
Predict the value of a response variable, informed by the value of an explanatory variable	Predict the value of a response variable, informed by the value of an explanatory variable
Assess whether the linear function is an appropriate summary description of the data, using residual plots	Assess whether the non-linear function is an appropriate summary description of the data, using residual plots

A data analyst might decide to use nonlinear fits for two reasons. First, inspections of a scatterplot and residual plot may indicate a clear non-linear pattern, one which could be more effectively summarized using a non-linear elementary mathematical function from algebra. As we will soon see, we have a variety of elementary functions to choose from. Second, the scientific community may have settled on the nature of the relationship between *x* and *y* as nonlinear, and the data analysis task is to estimate the parameters of the accepted function by finding the nonlinear best-fit curve. In the description to follow, it is convenient to separate non-linear fits into 2 general categories: those accomplished using polynomial functions, and those accomplished using what are known as transformations of variables.

Polynomial regression

In the article "Quantifying spatiotemporal overlap of Alaskan brown bears and people" (Journal of Wildlife Management [2005]: 810-817), investigators were concerned about human activity in the presence of foraging bears. Their specific concern was that sport fishing and boating might be displacing bears from sufficient access to salmon due to the presence of humans and their loud watercraft. Part of their research involved documenting the fishing activity of brown bears (*Ursus arctos*) through time.

Date (June 1 = 1)	Bear usage Bear-hr/day	Date (June 1 = 1)	Bear usage Bear-hr/day	Date (June 1 = 1)	Bear usage Bear-hr/day
11	11.3	24	11.3	40	22.0
12	15.1	25	18.4	41	26.0
13	6.6	27	16.2	44	10.5
14	12.9	28	19.5	45	18.6
15	12.1	31	35.8	46	21.1
17	18.1	32	37.1	49	11.9
18	20.9	33	45.7	50	13.7
19	17.6	36	34.8	51	13.7
20	11.0	37	25.6	54	6.3
21	24.6	38	26.7	55	1.8



The scatterplot at right displays the data on bear usage (bear-hours / day) vs. date (in days, 1 = June 1st) in 2003 at Wolverine Creek and Cove, Alaska. It is clear from the pattern of points that no linear best-fit line will do a reasonable job of describing the relationship between *x* and *y*. The points in the scatterplot appear to rise, level off near day 30 (June 30), and then fall as the days move through the month of July. The relationship between the amount of bear usage of Wolverine Creek and Cove and time is more complex than is captured by a linear relation.

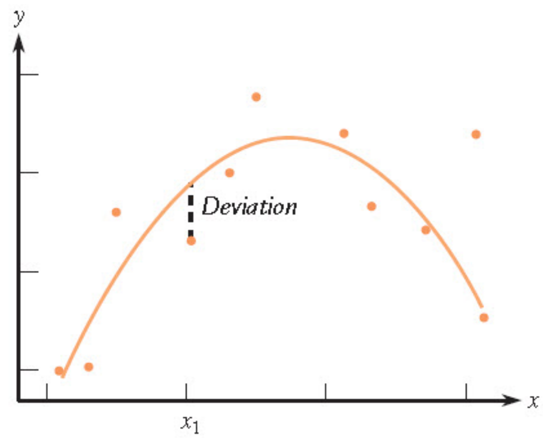
The usual interpretation of the slope of a best fit line is that as the explanatory variable changes by 1 unit, on average the response variable changes by a constant amount equal to the slope. These data do not exhibit a constant average increase in bear usage exists. Rather, it appears that the average changes in the *y* variable vary with *x*; the amount of change in *y* per unit change in *x* is a function, not a constant.

Quadratic functions exhibit this rise / level off / fall sort of appearance, and it would seem that a quadratic function of the form $\hat{y} = a + b_1 x + b_2 x^2$ is a more reasonable description of the pattern of points than a straight-line model. The values of the coefficients $a,\ b_1,$ and b_2 in this function can be determined to obtain a good fit to the data. (Note that the choice of the symbols for the coefficients is consistent with straight-line relationships, not with the typical algebraic description of a quadratic function, $y = f(x) = ax^2 + bx + c$.)

As is true of linear functions, algebra will enable one to initially interpret the graph and coefficients of a quadratic function:

- The sign of the coefficient of the quadratic term, b_2, indicates whether the quadratic curve opens up or down
- The maximum value of the quadratic function occurs where $x = -\dfrac{b_1}{2b_2}$
- The maximum value is estimated to be $y = f\left(-\dfrac{b_1}{2b_2}\right)$

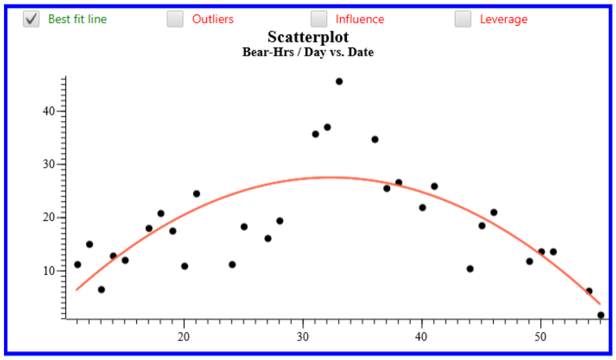
What are the best choices for the values of $a,\ b_1,$ and b_2? In fitting a line to data, we used the principle of least squares to guide our choice of slope and intercept. Least squares can be used to fit a quadratic function as well. The deviations , $y_i - \hat{y}_i$, are still represented by vertical directed distances in the scatterplot, but now they are vertical distances from the points to a parabola shown below. The task of the data analyst is to find values for the coefficients in the quadratic function so that the sum of squared deviations is as small as possible.



For a quadratic regression, the least squares estimates of *a*, b_1 and b_2 are those values that minimize the sum of squared deviations, $\sum (y - \hat{y})^2$, where $\hat{y} = a + b_1 x + b_2 x^2$.

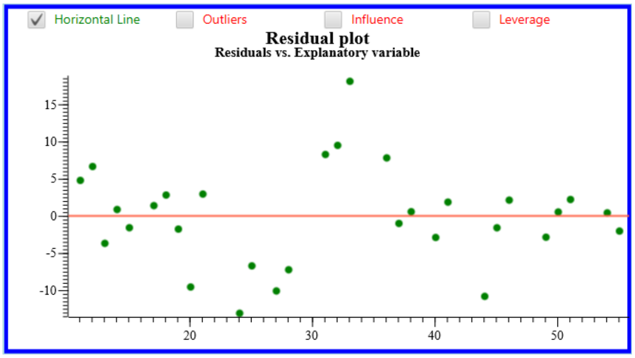
For quadratic regression, a measure that is useful for assessing fit is $R^2 = 1 - \frac{\text{SSResid}}{\text{SSTo}}$

where $SS\text{Resid} = \sum (y - \hat{y})^2$. Notice that the measure R^2 is defined in a manner analogous to r^2 for simple linear regression and is interpreted in a similar fashion. The notation r^2 is used only with linear regression to emphasize the relationship between r^2 and the correlation coefficient, r, in the straight-line fit. In nonlinear regression the symbol used is R^2. The formulas for computing the least-squares estimates in the general case are somewhat complicated without using matrices, so we will rely on statistical software packages to do the computations for us. Part of the output from fitting a quadratic regression to these data is shown at right.



The least squares coefficients are: $a = -20.9671$ $b_1 = 2.9958$ $b_2 = -0.0463$, and the least squares quadratic equation is: $\hat{y} = -20.9671 + 2.9958x - 0.0463x^2$.

If a least-squares line were fit to these data, it is not surprising that the line would not do a credible job of describing the relationship ($r^2 = 0.000001$). Both the scatterplot and the residual plot show a distinct curved pattern. The residual plot for the quadratic regression is shown below. Notice that there is no strong pattern in the residual plot for the quadratic case. For the quadratic regression, $R^2 = 0.556$ (as opposed to essentially zero for the least squares line). This means that 55.6% of the variability in the bear prevalence can be explained by an approximate quadratic relationship between bear prevalence and date of observation.

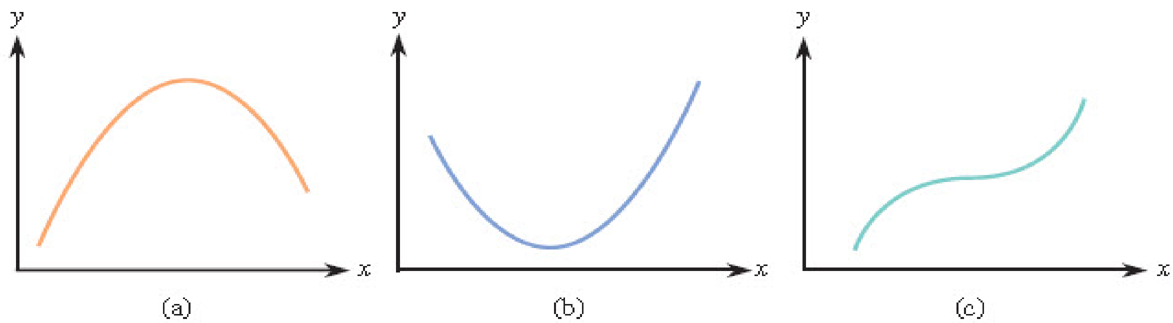


Linear regression and quadratic regression are special cases of polynomial regression. A least squares polynomial regression curve is described by a function of the form:

$$\hat{y} = a + b_1 x + b_2 x^2 + b_3 x^3 + \ldots + b_k x^k .$$

Recall that $p(x) = a + b_1x + b_2x^2 + b_3x^3 + ... + b_kx^k$ is referred to as a *k*th degree polynomial. The case of *k* = 1 results in linear regression ($\hat{y} = a + b_1x$) and *k* = 2 yields a quadratic regression ($\hat{y} = a + b_1x + b_2x^2$).

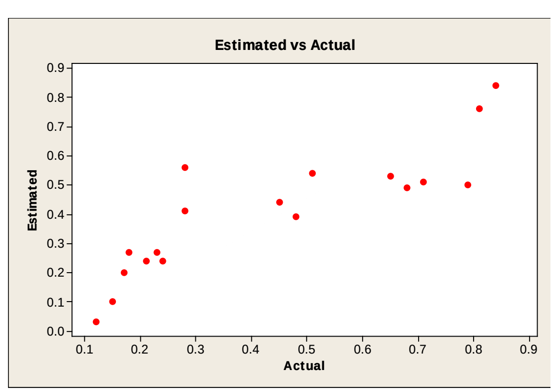
A less frequently (than quadratic) encountered special case is for *k* = 3, a cubic regression curve, $\hat{y} = a + b_1x + b_2x^2 + b_3x^3$. While quadratic curves have only a single bend, cubic curves tend to have two bends, as shown in (c) below.



A cubic fit was performed in the article, "Perceiving musical time" (Music Perception: An Interdisciplinary Journal [1990]:213-251). Twenty-three experienced music researchers and composers were asked to listen to a solo piano piece, comprised of 18 segments. The piece was described in the article as "…atonal with a pitch structure organized according to the principles of 12-note serialism…based on proportions derived from the Fibonacci series." Their data for the 18 segments are shown below:

Actual Location	Estimated Location	Actual Location	Estimated Location	Actual Location	Estimated Location
0.84	0.84	0.51	0.54	0.23	0.27
0.81	0.76	0.48	0.39	0.21	0.24
0.79	0.50	0.45	0.44	0.18	0.27
0.71	0.51	0.28	0.56	0.17	0.20
0.68	0.49	0.28	0.41	0.15	0.10
0.65	0.53	0.24	0.24	0.12	0.03

After listening to the piece twice, the musicians were given copies of different sections of the musical score and asked to locate the relative position of the segments in the piece. As an example, if the musician thought a section of music occurred three-fourths of the way through the piece, he or she would indicate 0.75. The "Estimated Location" is the median of the values given by the subjects in the study.

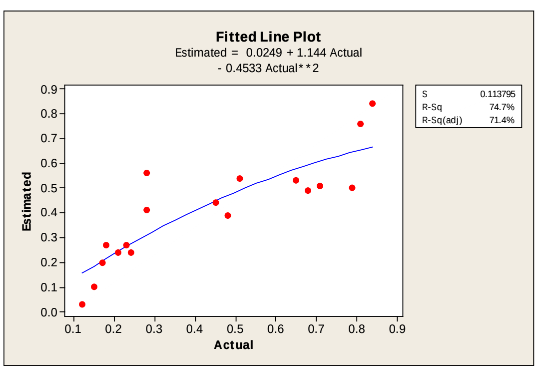


The relationship between x and y does not appear to be linear – it seems to have a bend in it. In the light of this, one might try using a quadratic regression to describe the relationship between the estimated and actual relative positions of the sections of the musical piece. Statistical software was used to fit a quadratic regression function and to compute the corresponding residuals. The least-squares quadratic regression is: $\hat{E} = 0.0244 + 1.1523A - 0.4657A^2$ Plots of the quadratic regression curve and the corresponding residual plot (below) have brought out a pattern we didn't notice in the scatterplot before. (This capability of residual plots to bring out the worst in graphs is one of the reasons we use them.) In this case the residual plot shows a curved pattern between the residuals and x – not something we like to see in a residual plot! Looking again at the scatterplot, we see that a cubic function might be a better choice than the quadratic function; assisted by the residual plot, we now see what appears to be two "bends" in the curved relationship – one in the neighborhood of x = 0.3 and another in the neighborhood of x = 0.7.

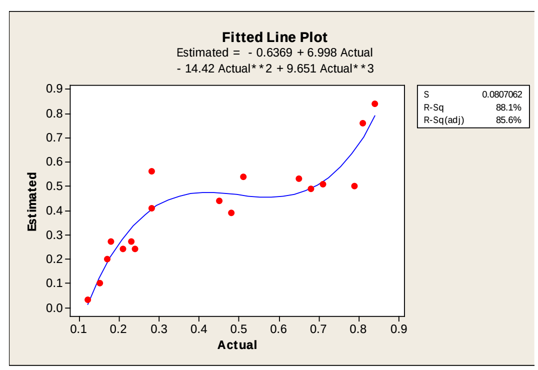
Computer software was used to fit a cubic regression, resulting in the curve shown in the quadratic and cubic fits. The cubic regression is:

$$\hat{E} = -0.6284 + 6.959A - 14.3823A^2 + 9.6594A^3 .$$

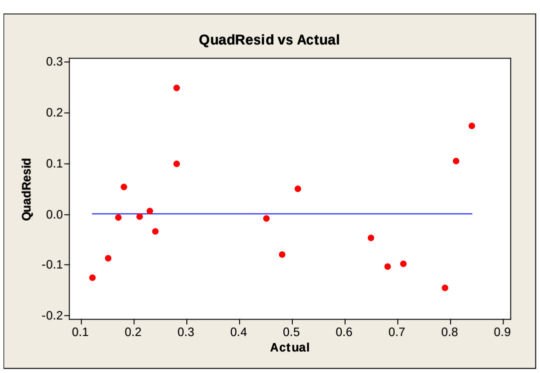
The cubic regression and the corresponding residual plots for the quadratic and cubic fits are shown below. The plots of the cubic fits do not reveal any troublesome patterns that would suggest we need to consider a choice other than cubic regression. And other good news is that R^2 has increased from 0.75 to 0.88, also suggesting the cubic fit is a better fit to the data than the quadratic fit.



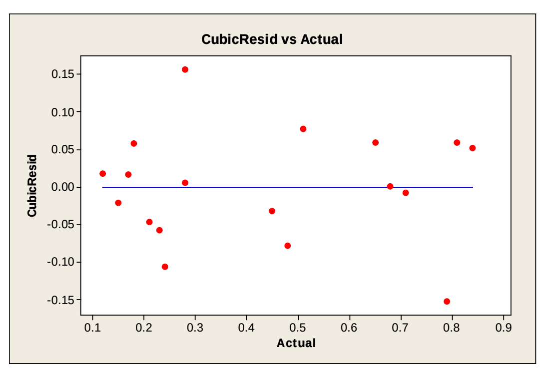
Quadratic fit



Cubic fit



Quadratic residual plot



Cubic residual plot

The investigators' inspection of the original scatterplot suggested to them that the subjects' judgments of the relative positions of the musical segments were fairly close to the actual positions at the beginning and end of the musical piece, but not so in the middle. They felt this might be due to a greater sense by the subjects of musical progress in the beginning and near the end of the piece, whereas the central part of the piece is "something of a mixture, where different ideas are combined and juxtaposed, so that the sense of goal-directed musical progress is weakened." [Sounds good to me!]

Transformations

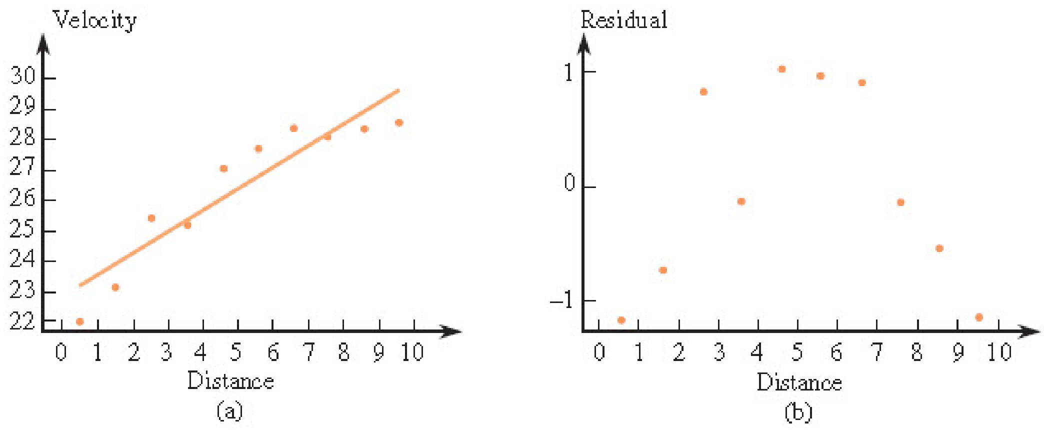
In general, our strategy for performing nonlinear fits using transformations is to find a way to legally change the x and/or y values so that a scatterplot of the transformed data has a linear appearance. A **transformation** (sometimes called a re-expression) involves using a simple function of a variable in place of the variable itself, to induce a linear relation between the transformed variables. For example, instead of trying to describe the relationship between x and y, it might be easier to describe the relationship between $\sqrt{x}$ and y or between x and log(y). And, if we can describe the relationship between, say, $\sqrt{x}$ and y, we will still be able to predict the value of y for a given x value. In addition, the interpretation of the slope is not only possible but reasonable. Common transformations involve taking square roots, logarithms, or reciprocals. To introduce you to the mechanics of using transformations, we will consider a square root transformation.

River Water Velocity and Distance from Shore

As fans of white-water rafting know, a river flows more slowly close to its banks (because of friction between the river bank and the water). To study the nature of the relationship between water velocity and the distance from the shore, data were gathered on velocity (in centimeters per second) of a river at different distances (in meters) from the bank. Suppose that the resulting data were as follows:

Distance	.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Velocity	22.00	23.18	25.48	25.25	27.15	27.83	28.49	28.18	28.50	28.63

A graph of the data exhibits a curved pattern, as seen in both the scatterplot and the residual plot from a linear fit.



Plots for the data
(a) scatterplot of the river data; (b) residual plot .

Let’s try transforming the *x* values by replacing each *x* value by its square root. We define

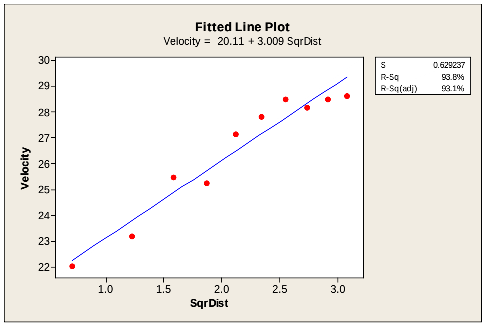
$$x' = \sqrt{x}$$

The resulting transformed data are given in Table 1 below.

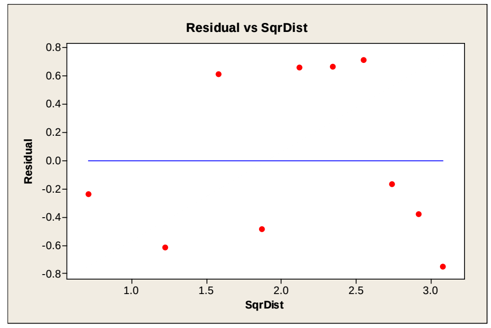
Original and transformed data of the river velocity

Original Data		**Transformed Data**	
x	***y***	***x'***	***y***
.5	22.00	0.7071	22.00
1.5	23.18	1.2247	23.18
2.5	25.48	1.5811	25.48
3.5	25.25	1.8708	25.25
4.5	27.15	2.1213	27.15
5.5	27.83	2.3452	27.83
6.5	28.49	2.5495	28.49
7.5	28.18	2.7386	28.18
8.5	28.50	2.9155	28.50
9.5	28.63	3.0822	28.63

The scatterplot (a) is of *y* versus *x′* (or equivalently *y vs.* $\sqrt{x}$). The pattern of points in this plot looks linear, and so we fit a least-squares line using the transformed data.



(a)



(b)

Plots for the transformed river data:
(a) scatterplot of *y* versus *x′*; (b) residual plot

Minitab output from this regression is shown below. The residual plot (b) shows no indication of a pattern. The resulting regression equation is: $\hat{y} = 20.1 + 3.01x'$ or, equivalently, $\hat{y} = 20.1 + 3.01\sqrt{x}$. The values of r^2 and s_e indicate that a straight line is a reasonable way to describe the relationship between *y* and *x′*. To predict velocity of the river at a distance of 9 meters from shore, we first compute $x' = \sqrt{x} = \sqrt{9} = 3$ and then use the sample regression line to obtain a prediction of *y*:

```
The regression equation is
Velocity = 20.1 + 3.01 SqrDist


Predictor      Coef   SE Coef      T      P
Constant    20.1102    0.6097  32.99  0.000
SqrDist      3.0085    0.2726  11.03  0.000


S = 0.629237   R-Sq = 93.8%   R-Sq(adj) = 93.1%
```

Figure 11

$\hat{y} = 20.1 + 3.01x' = 20.1 + 3.01(3) = 29.13.$

More Transformations

In the previous example, transforming the *x* values using the square root function worked well. We will now consider other transformations of variables. It is convenient to separate the transformations into two categories, for reasons that will become clear below:

- Situations where the explanatory variable (*x*) only is transformed
- Situations where the response variable (*y*) is transformed

While there is only one linear function, and only one description of the average change in *y* per unit change in *x* – constant – there are different patterns of values for nonlinear functions:

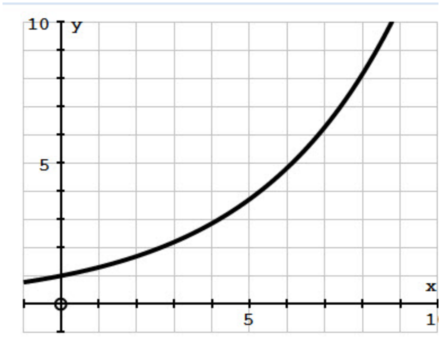
- There may be a single extreme (maximum or minimum) value of y. With increasing values of *x*, the expected values of *y* may rise then fall, or fall then rise.

- The increases in y per unit increase in *x* may be smaller for small values of *x*, or the increases in y per unit increase in x may be larger for small values of *x*.

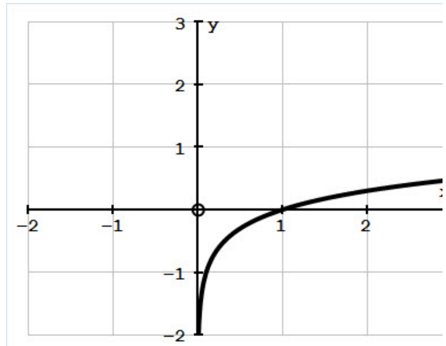
- The increases may be expressed in terms of proportions of *x* and/or *y*, not in units of *x* and/or *y*.

To help cope with this variety of ways data can be non-linear, there is a variety of functions we might try to fit to our data as we attempt to describe or explain the relationship between the variables. Some examples of elementary (second year algebra) functions, together with the associated transformations of variables leading to linearity, are collected below.

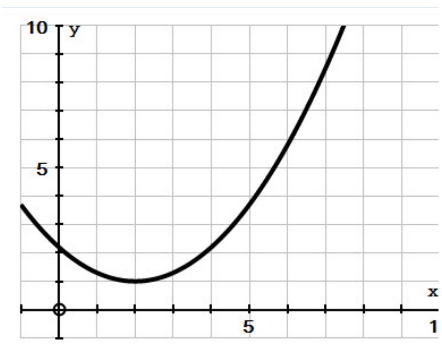
Functions (with associated transformations)



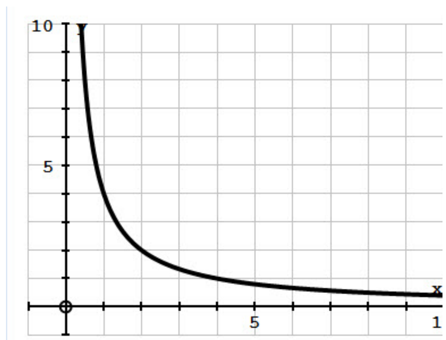
Exponential
$y' = \log(y)$



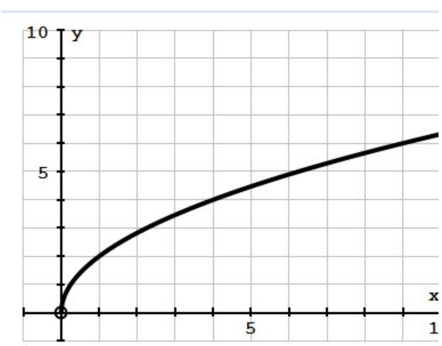
Logarithmic
$x' = \log(x)$



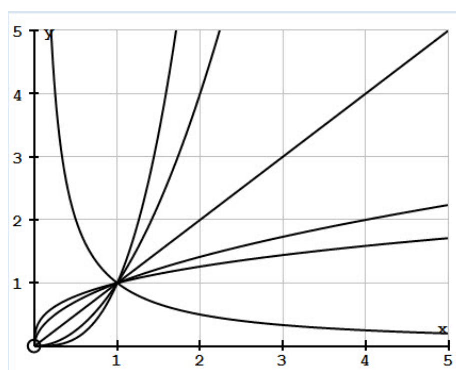
Quadratic
(No transformation)



Reciprocal
$x' = 1/x$



Square root
$x' = \sqrt{x}$



Power
$x' = \log(x)$; $y' = \log(y)$

The table below gives some guidance and summarizes some of the properties of the most commonly used transformations.

Commonly used transformations

Transformation	Mathematical Description	Try This Transformation if you believe that…
No transformation	$\hat{y} = a + bx$	The change in y is constant as x changes. A 1-unit increase in x is associated with, on average, an increase of b in the value of y.
Square root of x	$\hat{y} = a + b\sqrt{x}$	The change in y is not constant. A 1-unit increase in x is associated with smaller increases or decreases in y for larger x values.
Log of x*	$\hat{y} = a + b\log_{10}(x)$ or $\hat{y} = a + b\ln(x)$	The change in y is not constant. A 1-unit increase in x is associated with smaller increases or decreases in the value of y for larger x values.
Reciprocal of x	$\hat{y} = a + b\left(\frac{1}{x}\right)$	The change in y is not constant, as was true for the *log* function. Here, y has a limiting value of a as x increases, unlike the *log* function.
Log of y* (Exponential growth or decay)	$\widehat{\log(y)} = a + bx$ or $\widehat{\ln(y)} = a + bx$	The change in y associated with a 1-unit change in x is proportional to x.
Log of y* *and* log of x ("Power" function)	$\widehat{\log(y)} = a + b\log(x)$ or $\widehat{\ln(y)} = a + b\ln(x)$	The proportional change in y associated with a 1-unit change in x is proportional to x.

*The values of a and b in the regression equation will depend on whether $\log_{10}$ or ln is used, but the $\hat{y}'s$ and r^2 values will be identical. Notice that the two "log of y" transformations involve transforming the response variable.

The power transformation is a particularly interesting transformation. The power function, $y = ax^b$, is transformed to linearity by taking the logarithm (either common or natural) of both sides of the equation:

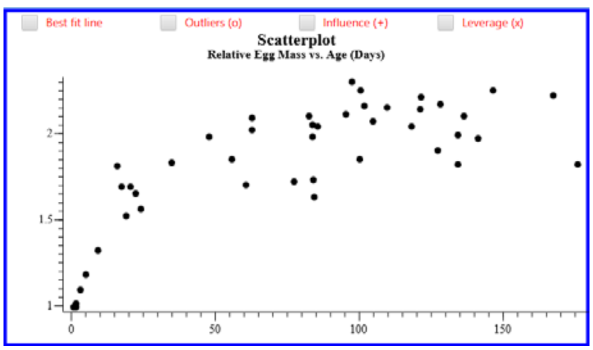
$$y = ax^b$$
$$\log(y) = \log(ax^b)$$
$$\log(y) = \log a + b\log\left(x\right)$$

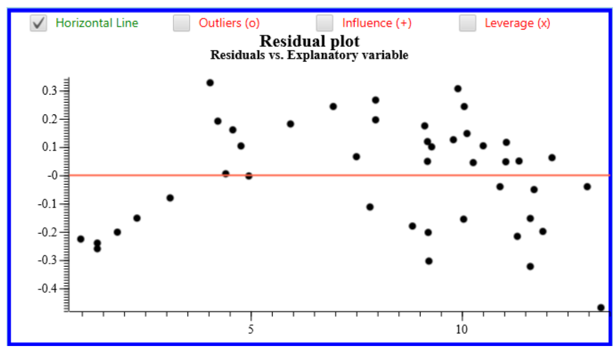
Thus, $x' = \log(x)$ and $y' = \log(y)$ result in a linear function, $y' = a + bx'$. What is interesting about the power function is that it includes raising to powers and taking roots. If *b* is a positive integer, a monomial results; if *b* is a fraction, such as one-half or one-third, the result is the same as taking a square root or cube root. In addition, if *b* is a negative integer, a reciprocal transformation is the result. The plots shown are for $y = x^b,\ b = {}^1\!/_3,\ {}^1\!/_2,\ 1,\ 2,\ 3,\text{ and } -1$.

Suppose that you have a batch of female veiled chameleons (*Chamaeleo calyptratus)* sitting around reproducing, as did Adams, Andrews, and Noble, described in their report, "Eggs under Pressure: Components of Water Potential of Chameleon Eggs during Incubation." (<u>Physiological and Biochemical Zoology</u>[2010]:207-214).

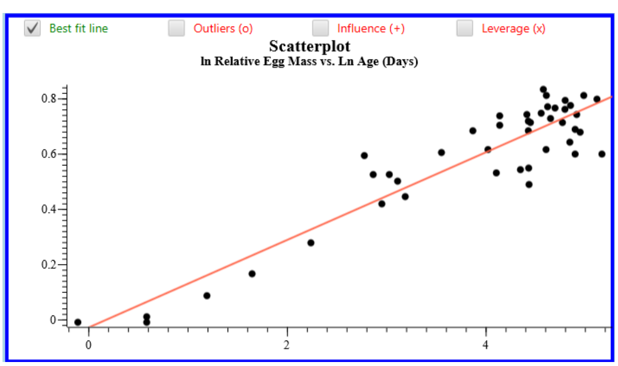
Of course, as you know, "embryos are diapausing gastrulae when eggs are laid and diapause persists several months." Further, realizing that "osmotic potential is generally assumed to dominate the net water potential of eggs, resistance of the eggshell to stretching also affects egg water potential," you decide to determine "osmotic potentials and pressure potentials" of the eggs. Skipping quite a bit of further discussion, you finally get data on the relative egg mass (Egg mass / Original Egg Mass) vs. Egg Age.

Age (days)	**Rel Egg Mass**	**Age (days)**	**Rel Egg Mass**	**Age (days)**	**Rel Egg Mass**
0.9	0.99	62.9	2.02	100.3	1.85
1.8	0.99	55.9	1.85	109.8	2.15
1.8	1.01	60.8	1.7	118.3	2.04
3.3	1.09	84.5	1.63	121.6	2.21
5.2	1.18	84.2	1.73	121.3	2.14
9.4	1.32	77.5	1.72	128.3	2.17
16.1	1.81	82.7	2.1	146.6	2.25
17.6	1.69	83.9	2.05	167.5	2.22
20.7	1.69	85.7	2.04	136.5	2.1
22.5	1.65	83.9	1.98	127.4	1.9
24.3	1.56	97.6	2.3	134.4	1.99
19.2	1.52	100.6	2.25	141.4	1.97
35	1.83	95.5	2.11	134.4	1.82
48	1.98	101.9	2.16	176	1.82
62.9	2.09	104.9	2.07		

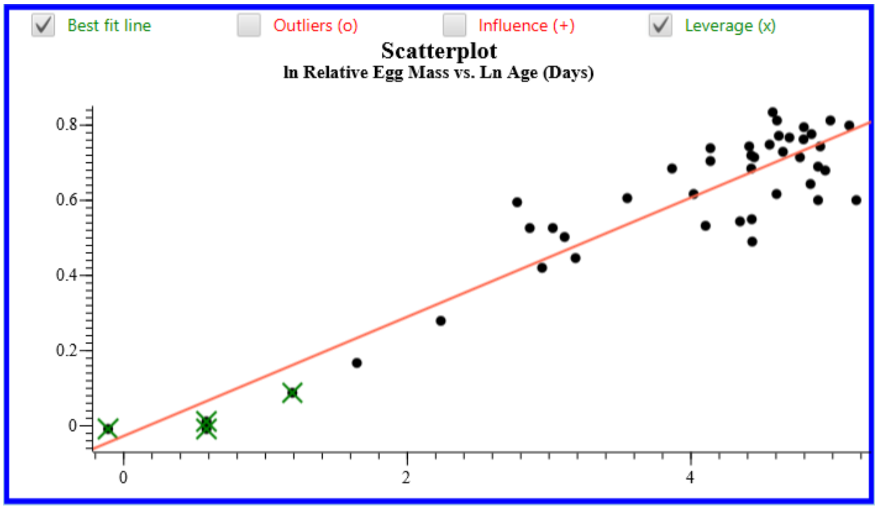




One's first impression might be that the data seems to fit a square root sort of function. However, after fitting $y = a + bx'$, where $x' = \sqrt{x}$, the residual plot exhibits an excessive amount of curvature. Still, the square root transformation seems reasonable. Perhaps a different power relationship will provide a better fit. The transformations to make this fit are: $x' = \ln(x)$ and $y' = \ln(y)$. Now, fitting $\widehat{y'} = a + \ln(x')$ (using natural logs this time) we see…



A closer look (with the aid of software!) reveals that there might be a problem with some high leverage points at the lower ages, but the points are very close to the best fit line, and in fact they are not overly influential in determining the fit.

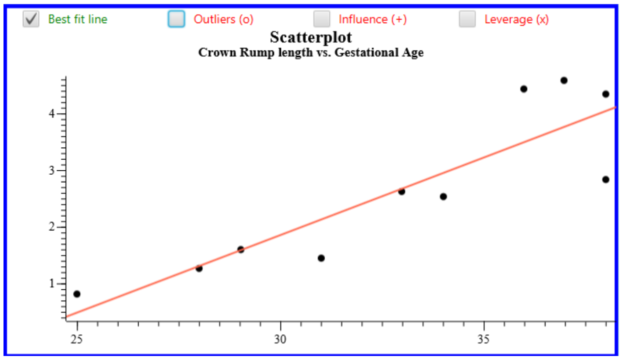


Suppose now that you have graduated from investigating chameleon eggs, and are now in Alaska, working with Mark McNay and his colleagues preparing a research report, "Diagnosing pregnancy, In Utero Litter Size, and Fetal Growth with Ultrasound in Wild, Free-Ranging Wolves." (Journal of Mammalogy [2006]:85-92).

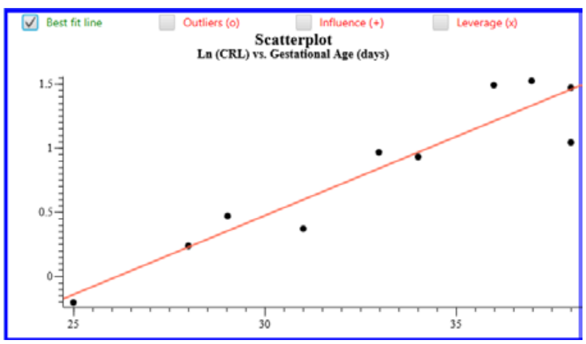
Did you get that part about Wild, Free-Ranging? Just so you know, here is what they look like when they are Wild and Free-Ranging and imagining you on a big plate. That's why one shoots darts loaded with 550 mg of tiletamine HCl and zolazepam HCl (available at Fort Dodge Lab, Fort Dodge, ***IOWA***) at them from the safety of a helicopter. Data on the fetal crown-rump length (distance from top of head to bottom of buttocks) in cm vs. fetal gestational age are shown below.

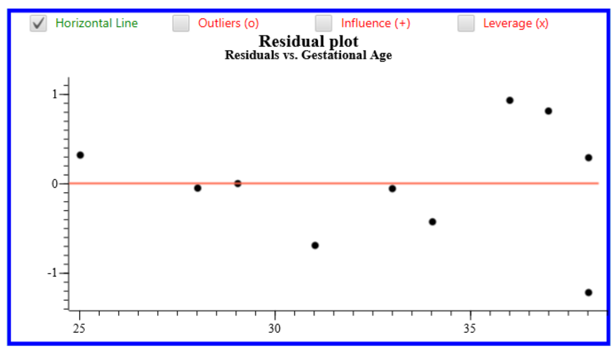
Gestational Age	CRL	Gestational Age	CRL
25	0.814	34.011	2.533
28.004	1.266	35.989	4.432
29.03	1.598	36.977	4.583
31.008	1.447	38.004	4.342
32.985	2.623	38.004	2.834

An initial plot of the data as well as theoretical biological considerations suggest that the actual relation between the crown rump length and gestational age is exponential, i.e. $y = ae^{bx}$.



Transforming the response variable with a logarithm gives us a best fit line, $\widehat{\ln y} = a + bx$, which in this case turns out to be: $\widehat{\ln(CRL)} = a + b(GA)$. The scatter plot and residual plot are presented below. (The transformation does not produce much change because the curvature was slight.)





Choosing a fit: combining the Best and the Brightest

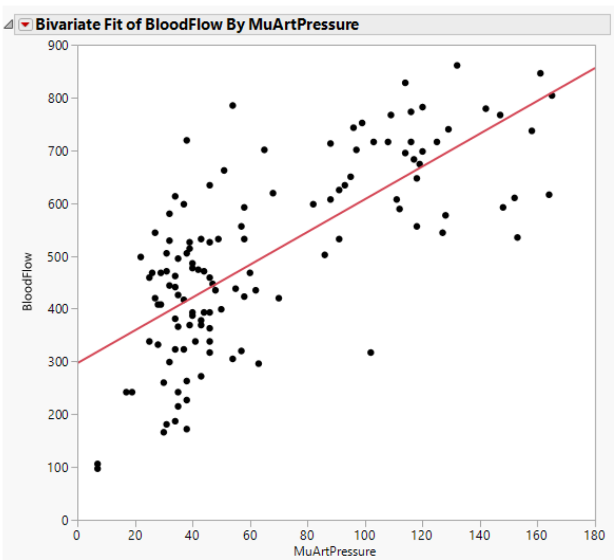
Once we reject the straight line as a plausible description of data, it is frequently the case that more than one of our polynomial or transformation strategies will produce a good fit to the data we have. Choosing a nonlinear regression procedure is a matter of statistical judgment guided by scientific wisdom. As scientists interested in explanation, we want to feel we have provided as complete an explanation, as maximal an accounting of the variability in *y*, as possible. This can sometimes translate into a ruthless search for a procedure – any procedure! – that has, as a result, small residuals and a large r^2.

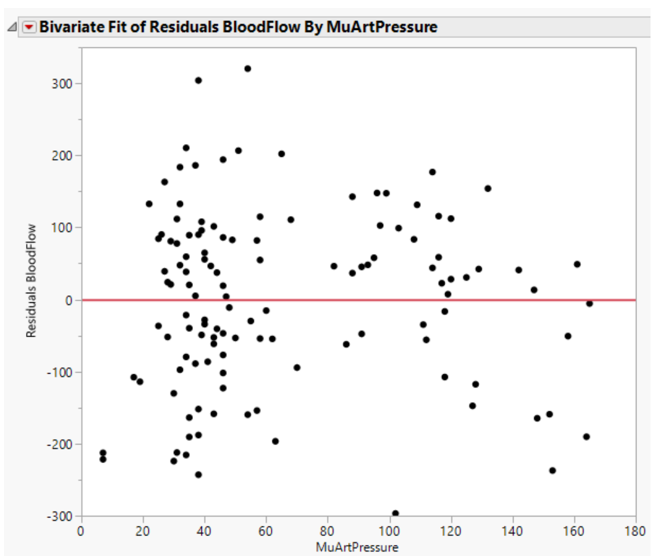
Unfortunately, there is a quick and easy (and wrong) way to do this by abusing polynomials!

Holding out for a hero: polynomials!?!?

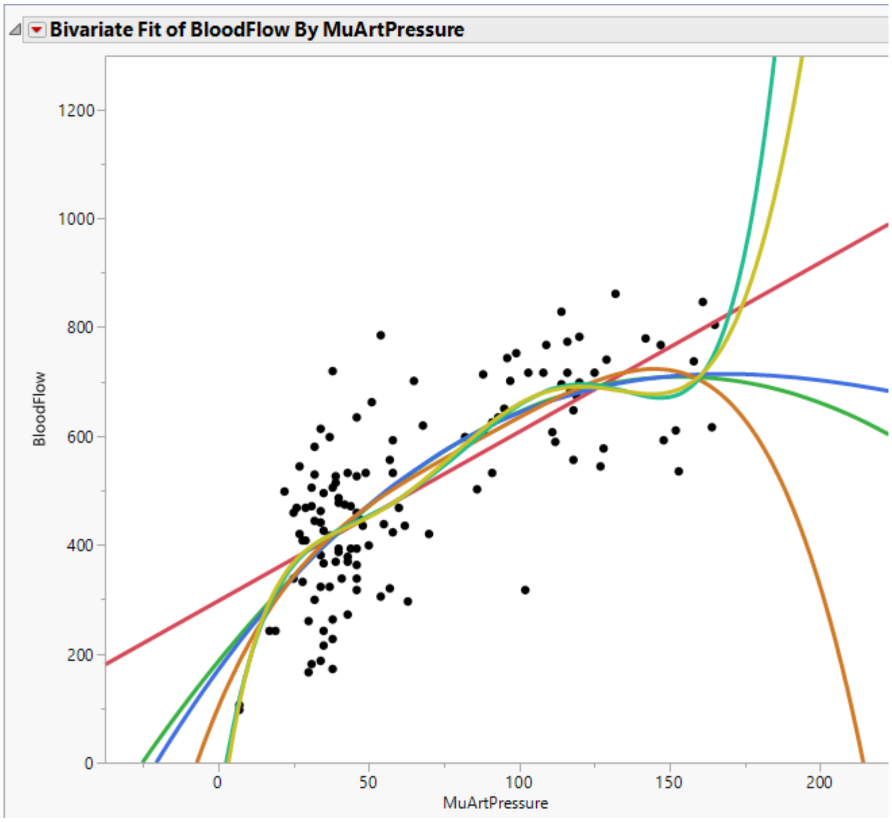
Polynomials are among the most useful and powerful functions in applied mathematics. Archimedes of Syracuse (287 – 212 BCE) is reported to have said that if he had a lever long enough and a fulcrum to use he could move the Earth. In the mathematical area of numerical analysis, it is well known that given a polynomial of high enough degree, any sufficiently tame function (i.e. differentiable) can be approximated to any desired accuracy using a polynomial. The higher the degree of the polynomial, the better the fit. However, in regression analysis this numerical analysis tool is more akin to the *deus ex machina* of Greek drama. (That is, a plot device whereby a seemingly unsolvable problem is suddenly and abruptly resolved by the inspired and unexpected intervention of some new event, character, ability or object. Thank you, Google!)

In regression analysis, it turns out that fits can be improved by the simple artifice of increasing the degree of the polynomial beyond the straight line (degree = 1) polynomial fit, $\hat{y} = a + bx$, to quadratics, cubics, quartics, quintics, and beyond. While this strategy may give a "best" fit, it will not in general give the "brightest" fit. My meaning of "brightest fit" here is the fit that mirrors the underlying physical, chemical, biological, psychological, etc. relationship or process under investigation. Only very rarely would a relationship be a polynomial beyond quadratic degree. (The one we saw above is the first one seen in 30 years of looking!) Bottom line: if you are seeking explanation or understanding – as distinguished from mere prediction – high degree polynomials seldom if ever rise to the level of good nonsense! To illustrate and hopefully engender future rejection of the siren song of higher degree polynomials, please consider an experiment the subject of which was the relationship between blood flow and blood pressure in rats. [Carreira, et al. (2014). Diaphragmatic function is preserved during severe hemorrhagic shock in the rat. Anesthesiology 120:425-35.] The linear fit and associated residual plot are shown below, using the way cool capabilities of the JMP statistics program.





For the linear fit $r^2 = 0.518148$ (all those decimals just show our sense of humor, but bear with me here), and the plots indicate the presence of curvature. Suppose now a data analyst decides to ruthlessly pursue a larger r^2 (technically now, a larger R^2) by fitting higher and higher degree polynomials. Here are polynomial fits up to sixth degree, using JMP:



Notice the [alleged] Good, the Bad, and the Ugly:

Degree of polynomial	r^2
1	0.518148
2	0.551057
3	0.551284
4	0.554082
5	0.568641
6	0.569111

[alleged] Good:

The r^2 increases with the degree of the polynomial.

Bad:

Outside the range of experimental values of mean arterial pressures, the polynomials seriously disagree about the nature of the relationship, and thus it is difficult to believe that this class of functions has captured the nature of the relationship between blood flow and blood pressure.

Ugly:

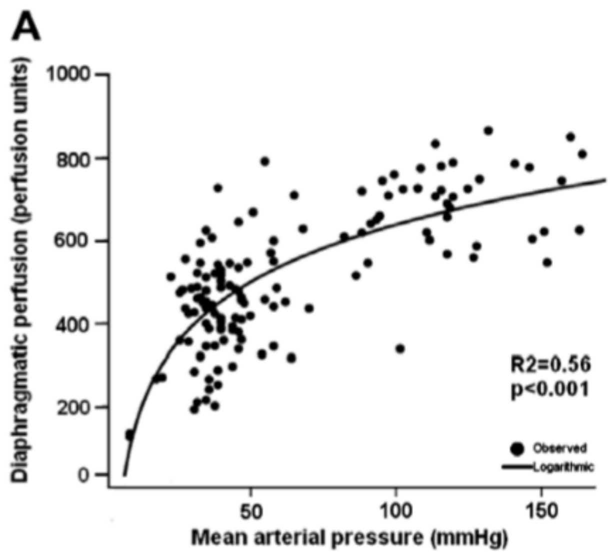
How would one possibly interpret the polynomial coefficients???

Carreira, et al. explained their results in the extraordinarily opaque (to this writer) doctor lingo, but suggested a very simple explanatory nonlinear relation:

MAP = mean arterial pressure

$BloodFlow = \alpha + \beta \ln\left(MAP\right) + \varepsilon$

The associated differential equation indicates this elementary and easily interpretable simplicity: $\frac{dF}{dP} = \frac{k}{P}$. Here is their plot of the data – notice also the respectable r^2:

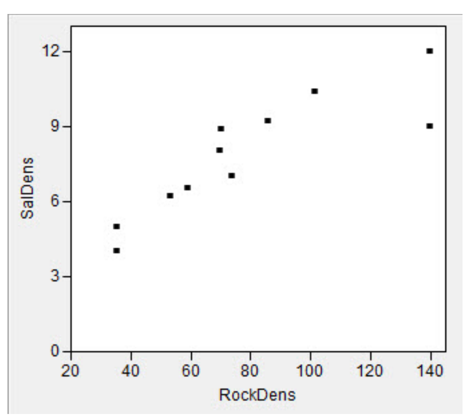


The message here is that while increasing the degree of a polynomial increases the proportion of variability (allegedly) "explained by" the polynomial, higher degrees do not result in a better understanding of the underlying relation or process under investigation. The increase in r^2/R^2 is a numerical chimerical illusion.

Holding out for a hero: the scientific received view

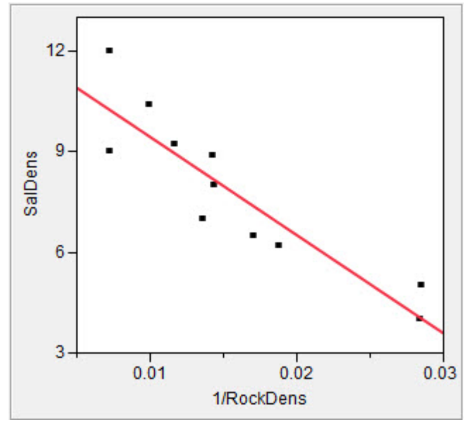
Picking the "Brightest" fit – the one in agreement with accepted scientific laws – would usually be the preferred strategy.

Frequently, regression is used in the tentative creation of scientific laws. In cases such as these an investigator may reason from her knowledge of science and be able to reject some possible nonlinear functions in favor of others. We found an interesting example of scientific judgment in an experimental study of factors that



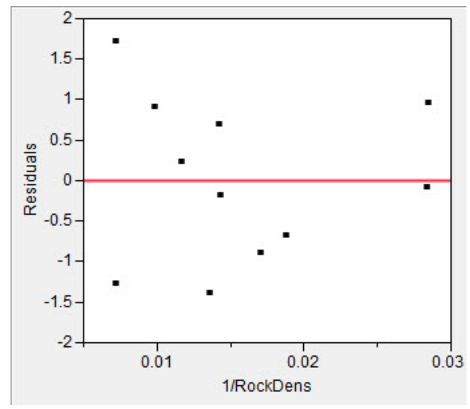
Sal Density vs. Rock Density

influence population density of salamanders in the paper, “The relationship between rock density and salamander density in a mountain stream” (Herpetologica [1987]: 357-361). The investigators created a range of habitats for salamanders (*Desmognathus quadramaculatus*) by placing different sized rocks and pebbles in a small stream in the Southern Appalachian Mountains. Three months later they returned to measure the population density of the salamanders. The scatterplot of their original data are shown at right. The densities are measured in count / 1.4 square meters.



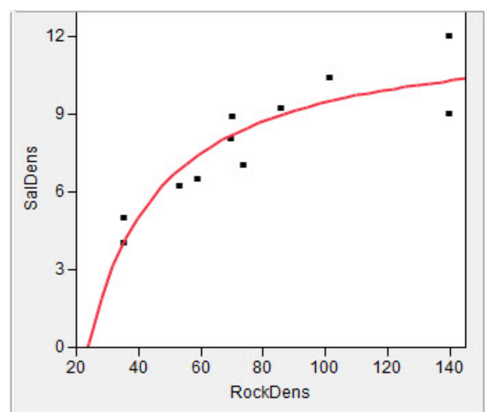
SDensity vs. 1/RDensity

Inspection of the potential elementary functions suggests more than one plausible function to use to fit the data. The researchers chose a reciprocal regression function, $\hat{y} = a + b\left(\frac{1}{x}\right)$, for two reasons: (1) it was the best, and (2) it was the brightest. The r^2 was greatest for the reciprocal function, which was icing on the cake. More important, the reciprocal function made sense scientifically. The investigators felt that there would be an upper limit to the population density since the stream bed is a nonrenewable resource and the stream therefore had a limit in the number of salamanders that could be sustained.



Residual Plot

This limit is known as the “carrying capacity” of an environment and is estimated by the value of the intercept, *a*, in the regression function. The choice of a reciprocal transformation may seem odd to you because the reciprocal transformation shown above is the only one that is falling with increasing *x* values. Remember, though, that functions are transformed into mirror images by $f(-x)$ and $-f(x)$.



Original scale

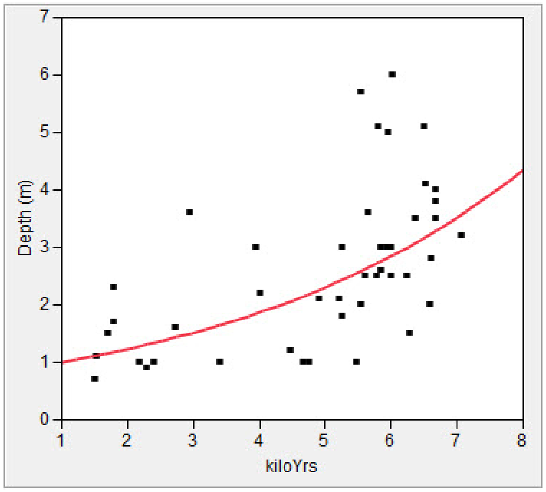
After defining the transformation $x' = \frac{1}{x}$ and fitting the resulting data, the best fit line was calculated using JMP. The best fit line, ($\hat{y} = 12.37 - 292.6x'$, $r^2 = 0.82$) and the residual plot are shown above. The scatterplot of the original (untransformed) data with the best fit regression equation superimposed is shown at right.

Here is an example of a nonlinear fit, this one in the absence of settled scientific theory, from the article in "Sea-Level Rise on Eastern China's Yangtze Delta" (Journal of Coastal Research [1998]: 360-366).

The researchers used pollen and microfossil records in radiocarbon-dated samples of peat from core samples as well as archeological data to produce historic water levels in the Yangtze delta of China to study the pattern of the rising of sea level. Geologic and hydrologic data are notorious for not having the benefit of common scientific models (i.e. there is no Brightest), and the researchers elected to fit an exponential model to their data as the Best summary of the relation between sea-level and time. Their data are reproduced in the table below and are relative to the present sea-level and present time. The variable "Kilo-Years BP" is thousands of years before the present; the depth variable is the depth compared to the current sea level. As an example, based on the measurements available the researchers inferred that 7,064 years ago the sea-level was 3.2 meters below the current level. The "Log of Depth" is the common (base 10) logarithm of the Depth. A scatterplot of the Depth vs. Kilo-Years BP with a fitted exponential function is shown.

Kilo-years BP	Depth (m)	Log of Depth	Kilo-Years BP	Depth (m)	Log of Depth	Kilo-Years BP	Depth (m)	Log of Depth
7.064	3.2	0.5051	5.930	3.0	0.4771	4.660	1.0	0.0000
6.680	4.0	0.6021	5.845	2.6	0.4150	4.470	1.2	0.0792
6.670	3.8	0.5798	5.845	3.0	0.4771	4.000	2.2	0.3424
6.670	3.5	0.5441	5.790	5.1	0.7076	3.950	3.0	0.4771
6.600	2.8	0.4472	5.780	2.5	0.3979	3.407	1.0	0.0000
6.580	2.0	0.3010	5.640	3.6	0.5563	2.950	3.6	0.5563
6.510	4.1	0.6128	5.600	2.5	0.3979	2.720	1.6	0.2041
6.500	5.1	0.7076	5.530	5.7	0.7559	2.393	1.0	0.0000
6.365	3.5	0.5441	5.530	2.0	0.3010	2.285	0.9	-0.0458
6.275	1.5	0.1761	5.470	1.0	0.0000	2.180	1.0	0.0000
6.227	2.5	0.3979	5.260	3.0	0.4771	1.790	2.3	0.3617
6.008	6.0	0.7782	5.260	1.8	0.2553	1.780	1.7	0.2304
6.000	2.5	0.3979	5.210	2.1	0.3222	1.691	1.5	0.1761
6.000	3.0	0.4771	4.901	2.1	0.3222	1.530	1.1	0.0414
5.960	5.0	0.6990	4.750	1.0	0.0000	1.510	0.7	-0.1549

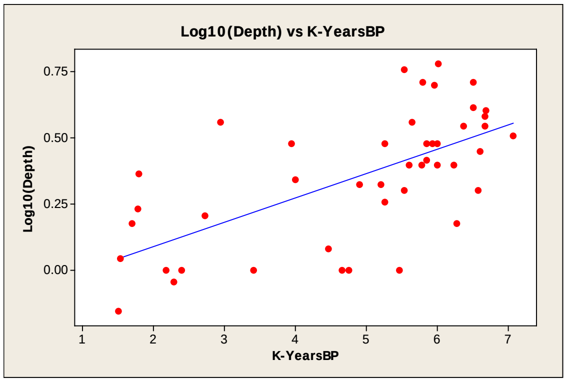
The scatterplot is typical of data seen when two variables are related by an exponential function. The change in *y* for increasing values of *x* is less for small *x* values than for large values of *x*. For these data, think in changes in *x* of units of 1000 years. Another feature common to exponential relations is that the variability about the line is greater for larger values of *x* than it is for smaller values of *x*.



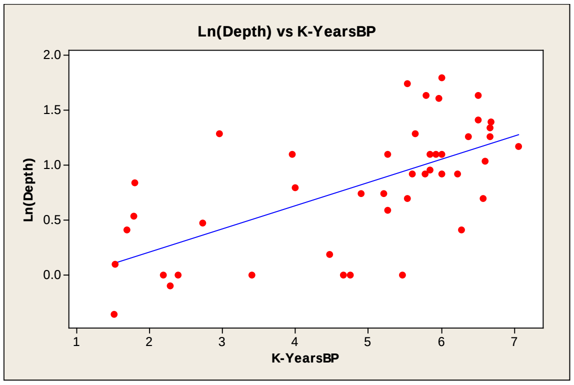
Sea-level vs. time

The plot of sea level vs. time hints that using logarithms and transforming the *y* variable (the depth) will be in order. Two standard logarithmic functions are commonly used for such transformations – the common logarithm (log base 10, denoted by log or $\log_{10}$) and the natural logarithm (log base e, usually denoted by ln, but sometimes as $\log_e$). Either the common or natural log can be used; the only difference in the resulting scatterplots is the scale of the transformed *y* variable. This can be seen below where the scatterplots of y' vs. *x* for both logarithmic transformations are shown, together with the best fit lines. These two scatterplots show the same pattern.

The resulting regression equations using the common log transformation are $\widehat{y}' = -0.093 + 0.0915K$, or equivalently, $\widehat{\log(y)} = -0.093 + 0.0915K$. For the natural log transformation the resulting regression equation is $\widehat{y}' = -0.215 + 0.2106K$, or equivalently $\widehat{\ln(y)} = -0.215 + 0.2106K$.



$y' = \log(x)$



$y' = \ln(x)$

Recovering a curve after using transformations

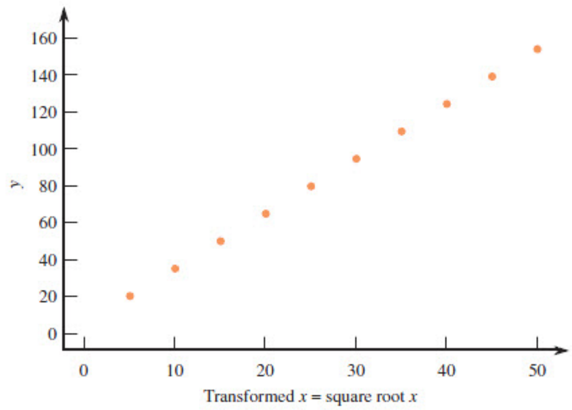
The objective of a regression analysis is usually to describe the approximate relationship between *x* and *y* with an equation of the form *y* = some function of *x*. If we have transformed only *x*, fitting a least-squares line to the transformed data results in an equation of the desired form, for example,

$$\hat{y} = 5 + 3x' = 5 + 3\sqrt{x}, \text{ where } x' = \sqrt{x}$$

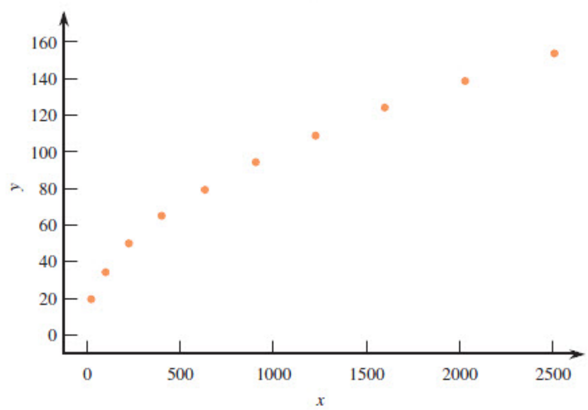
or

$$\hat{y} = 4 + 0.2x' = 4 + 0.2\frac{1}{x}, \text{ where } x' = \frac{1}{x}.$$

These functions specify lines when graphed using *y* vs. x', and they specify curves when graphed using *y* vs. *x*.



(a)



(b)

If the *y* values have been transformed, after obtaining the least-squares line the transformation can be "back transformed" to yield an expression of the form *y* = some function of *x* (as opposed to y' = some function of *x*). For example, to reverse a logarithmic transformation ($\widehat{y'} = \log(y)$), we can take the antilogarithm of each side of the equation. To reverse a square root transformation ($\widehat{y'} = \sqrt{y}$), we can square both sides of the equation, and to reverse a reciprocal transformation ($\widehat{y'} = \frac{1}{y}$), we can take the reciprocal of each side of the equation.

For the common log transformation used with the sea-level data, $\widehat{y'} = \log(y)$ and the least-squares line relating y' and *x* was $\widehat{y'} = -0.093 + 0.0915K$ or equivalently, $\widehat{\log(y)} = -0.093 + 0.0915K$. To reverse this transformation, we take the antilog of both sides of the equation:

$$10^{\log(y)} = 10^{-0.093+0.0915K}$$

Using properties of logs and exponents we know that

$$10^{\log(y)} = y \quad \text{and} \quad 10^{-0.093+0.0915K} = \left(10^{-0.093}\right)\left(10^{0.0915K}\right)$$

Finally we get

$$\hat{y} = \left(10^{-0.093}\right)\left(10^{0.0915K}\right) = 0.8072\left(10^{0.0915K}\right) = 0.8072\left(1.233\right)^{K}$$

This equation can now be used to predict the *y* value (sea-level) for a given *x* (thousands of years ago). For example, the predicted sea-level 2500 years ago (K = 2.5) is:

$$\hat{y} = 0.8072\left(1.233\right)^{K} = 0.8072\left(1.233\right)^{K} = \left(0.8072\right)\left(1.6934\right) = 1.3669$$

A final warning about "back transformations"

The process of transforming data, fitting a line to the transformed data, and then undoing the transformation to get an equation for a curved relationship between *x* and *y* usually results in a curve that provides a reasonable fit to the sample data, but it is not the least-squares curve for the data. For example, we used a transformation to fit the curve $\hat{y} = \left(10^{-0.093}\right)\left(1.233\right)^{K}$ above. However, there may be another equation of the form $\hat{y} = a\left(10\right)^{bx}$ that has a smaller sum of squared residuals for the *original* data than the one obtained using transformations. Finding the least-squares estimates for *a* and *b* in an equation of this form is mathematically complicated. Fortunately, the curves found using transformations usually provide reasonable predictions of *y* even in the original scales.