Linear Regression and Linearizing Data

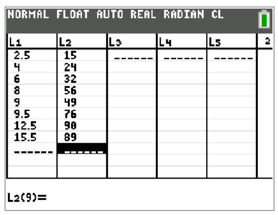
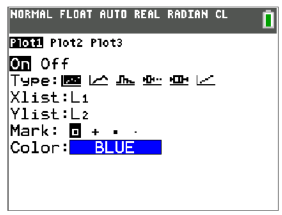
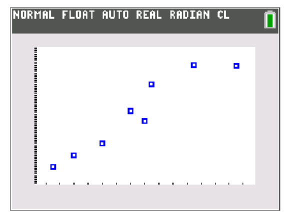
Linear Regression

- Linear Regression is a numerical process that fits a "best line" to a set of data. The process minimizes the square of the distance from the best fit line to each data point. A quantitative measure of how well the line fits the data is given by the *correlation coefficient*. The closer the correlation is to 1 or −1, the better the fit; closer to 0 is a bad fit.

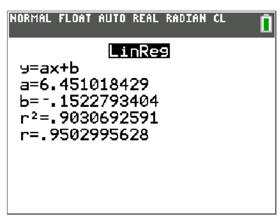
Example 1 The following data summarizes the width and age of a certain variety of oak tree in a forest.

Diameter (in.)	2.5	4.0	6.0	8.0	9.0	9.5	12.5	15.5
Age (years)	15	24	32	56	49	76	90	89

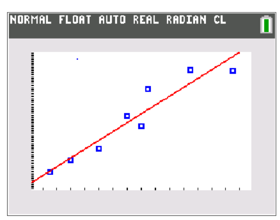
(a) Make a scatterplot of the data. Press: **STAT** **1:EDIT,** and enter Diameter into L_1 and Age into L_2. Next, press **2 ND** **Y =** (STAT PLOT) **1** and set up Plot1. Lastly, to make the scatterplot, press **ZOOM** **9:ZoomStat**.

(b) Find the line of best fit. Press **STAT** **▸** and select **4:LinReg (ax+b)** and press **ENTER**. Note, if you do not see r^2 and r, press **MODE** and set Stat Diagnostics to **ON.**



(c) Enter the equation into the equation editor **Y =** and press **GRAPH**.



(d) Use your linear model to estimate the age of an oak tree with a circumference of 54 inches.

Example 2 Biologists have observed that the chirping rate of a cricket of a certain species appears related to temperature. The following data summarizes the chirping rate for various temperatures.

Temperature (°F)	50	55	60	65	70	75	80	85	90
chirps/min	20	46	79	91	113	140	173	198	211

(a) Make a scatterplot of the data.
(b) Find and graph the regression line (line of best fit).
(c) Use the linear model to predict the chirping rate at 100°F.

Transforming Non-Linear Data to Linear

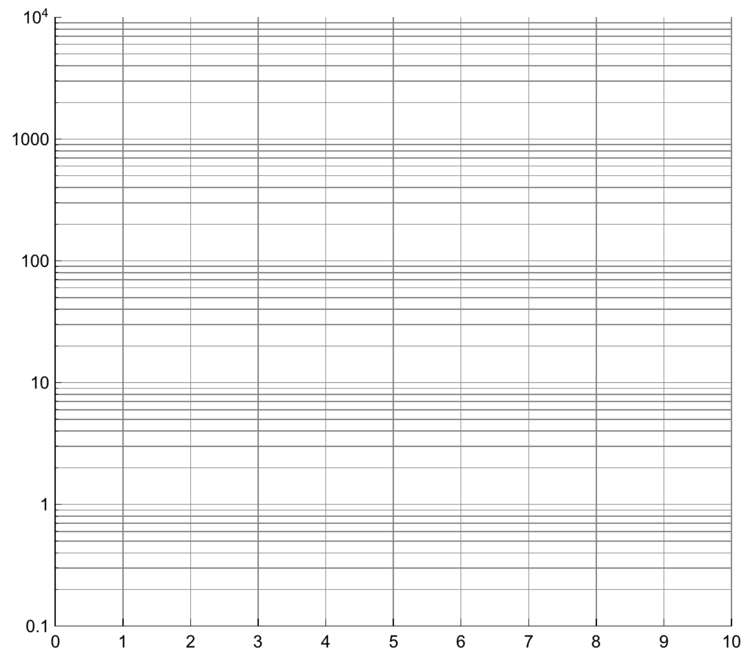
One use of logarithms is to transform data that exhibit either a power trend or exponential trend. Your calculator probably has both a power regression and exponential regression feature, but it is actually transforming the data to linear, finding the line of best fit, and then *un-transforming* the linear model back to the original form.

Transforming Exponential Data to Linear

Suppose we have data that we presume exhibits an exponential trend in the form of $y = a\,b^x$. The data can be *linearized* by taking the Log of the y values, that is transforming a data point (x, y) to $(x, \text{Log}(y))$. This can be done by hand by plotting the original data on **Log-Linear** graph grid.

Example 3 Graph the following data on the Log-Linear graph grid.

x	0	1	2	3	4	5	6
y	5	13	34	88	230	595	1545



Why Linearizing Works

Suppose we begin with the equation $y = a\,b^x$. Take the Log of each side and simplify:

$$\begin{aligned}\text{Log}(y) &= \text{Log}(a\,b^x)\\ &= \text{Log}(a) + \text{Log}(b^x)\\ &= \text{Log}(a) + x\,\text{Log}(b)\end{aligned}$$

Substituting $Y = \text{Log}(y)$, $A = \text{Log}(a)$, and $B = \text{Log}(b)$, we have

$$Y = A + B\,x$$

which is a linear function with slope B, and y-intercept A. Note that the y variable has been "Logged", but not the x variable. Once the linear equation is found, we can recover the exponential model, $y = a\,b^x$, by noting $a = 10^A$ and $b = 10^B$.

To linearize the data with your calculator, enter the original data into lists L_1 and L_2. To transform the data, move your cursor to the L_3 heading and enter Log(L_2). Next, set up **Plot1** using L_1 and L_3, and find the regression line using **4:LinReg (ax+b) L_1,L_3**. Note, to get **L_1**, press 2 ND 1; similar for **L_2**, etc.

(a) Find the slope and intercept for the regression line for the linearized data.
(b) Un-transform the linear model back to exponential.
(c) Compare the exponential model from step (b) to the calculators' model using **0:ExpReg**.

Transforming Power Data to Linear

Transforming data that appears of power form, i.e, in the form $y = a\,x^b$, is similar to the transformation of exponential data except, **both** the x and y values are logged. To see why, consider the power model, $y = a\,x^b$. When the equation is Logged, we get

$$\begin{aligned} y &= a\,x^b \\ \text{Log}(y) &= \text{Log}\left(a\,x^b\right) \\ \text{Log}(y) &= \text{Log}(x) + \text{Log}\left(x^b\right) \\ \text{Log}(y) &= \text{Log}(a) + b\,\text{Log}(x) \end{aligned}$$

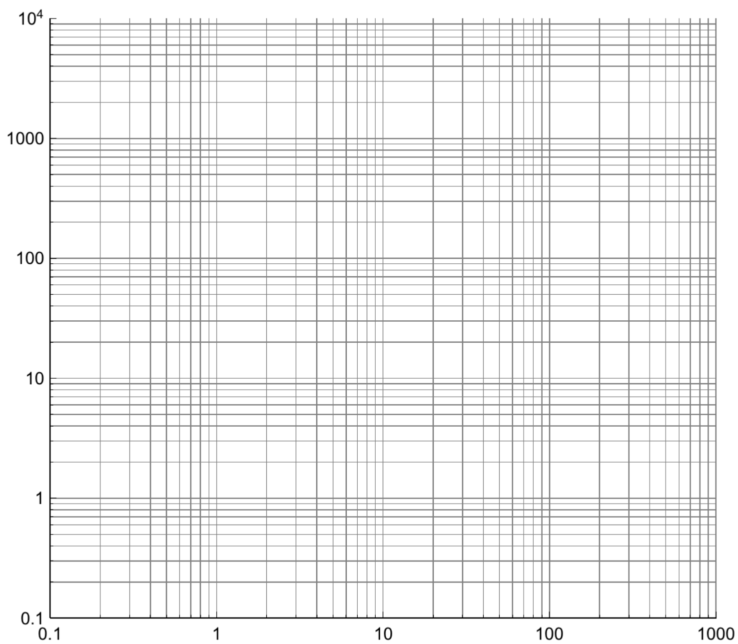
And, using the substitutions $Y = \text{Log}(y)$, $A = \text{Log}(a)$, and $X = \text{Log}(x)$, we obtain the linearization

$$Y = A + b\,X$$

again, a linear equation with A the y-intercept, and b the slope. This time, you can see both the x and the y variables are transformed, that is $(x,\, y) \to (\text{Log}(x),\, \text{Log}(y))$. To linearize the data with your calculator, let $L_3 = \text{Log}(L_1)$, and $L_4 = \text{Log}(L_2)$. Find the linear regression using L_3 and L_4, and un-transform the linear model with $a = 10^A$ and $b = b$.

Example 4 Graph the following data on the Log-Log graph grid.

x	2	5	10	40	100	300	700
y	5.9	18.7	44.5	251.5	791	3121	9001



(a) Linearize the data by entering the original data into L_1 and L_2, and the transformed logged data into L_3 and L_4.
(c) Un-transform the linear data by noting the slope is b, and $a = 10^A$.
(d) Compare your power model with the power model obtained from your calculator using **A:PwrReg**.
(e) Use your power model to predict y, (typically notated $\hat{y}$) when $x = 25$.

Example 5 A student is linearizing exponential data and obtained the linear regression line $y = 1.045\,x + 1.19$. Find the associate exponential function.

Example 6 A student is linearizing power data and obtained the linear regression line $y = 0.34\,x + 2.32$. Find the associated power function.

Example 7 Use your calculators **PwrReg** operation to find a power model for the following data. Explain the result, and a remedy.

x	1	2	3	4	5
y	0	9	24	45	72

Example 8 As sunlight passes through the waters of lakes and oceans, the light is absorbed and the deeper it penetrates, the more its intensity diminishes. The function that models the light intensity *I* at a depth *x* is called the Beer-Lambert Law. A photometer was used to investigate the light penetration in a northern lake, obtaining the following data:

Depth x (ft)	5	10	15	20	25	30	35	40
Light Intensity I (lm)	13.0	7.6	4.5	2.7	1.8	1.1	0.5	0.3

(a) Transform the data for both exponential and power and determine graphically the best model.
(b) Find the regression line for the model, and un-transform the model back to the original data.
(c) If the light intensity drops below 0.15 lumen (lm) a certain species of algae will not grow. Find this crucial depth.
(d) If the model found in (b) is exponential convert it to the form $I = I_o e^{-kx}$.

Example 9 The following data can be modeled by a logarithmic model in the form $y = a + b\ln(x)$.

x	61	610	1037	2074	3050	4087	5002	6100	7015
y	0.9742	1.1671	1.1988	1.2819	1.3074	1.3202	1.3549	1.3694	1.3751

(a) Plot the original data.
(b) Plot the transformed data by taking the natural log of the *x* values.
(c) Find and compare the logarithmic model using the data in (a) and your **9:LnReg** function on your calculator, with the linear model using the transformed data in part (b). What do you notice?