LEAST SQUARES NONLINEAR REGRESSION ANALYSIS

In the last unit we found a method for fitting a straight line to a set of (x, y) data pairs. This is OK if the data are expected to fit a linear theoretical model. In engineering we often want to fit experimental data to nonlinear models (equations).

One way we can build on our previous linear expertise is to TRANSFORM the nonlinear equation such that it becomes a linear one. As an example, suppose we wish to fit experimental data to the equation:

$$V = Ge^{-ct}$$

where V is a measured voltage at time t. In order to fit the equation we need to find the "best" values for parameters G and c. For this particular example we could *LINEARLY TRANSFORM* the equation by taking natural logarithms of both sides of the equation to get:

$$\ln(V) = \ln(G) - ct$$

This transformed equation now looks like a linear one with the y-values being replaced with ln(V) values. In this case the x-values will just be the t values. The intercept of the best-fit straight line would be ln(G) and the slope would be -c.

$$\left.\begin{matrix}\ln(V)\\ y\end{matrix}\right\} = \left\{\begin{matrix}\ln(G)\\ b\end{matrix}\right. + \left\{\begin{matrix}-ct\\ ax\end{matrix}\right.$$

EXAMPLE: The following voltages were measured at the specified times. Fit the above equation and find the parameters G and c.

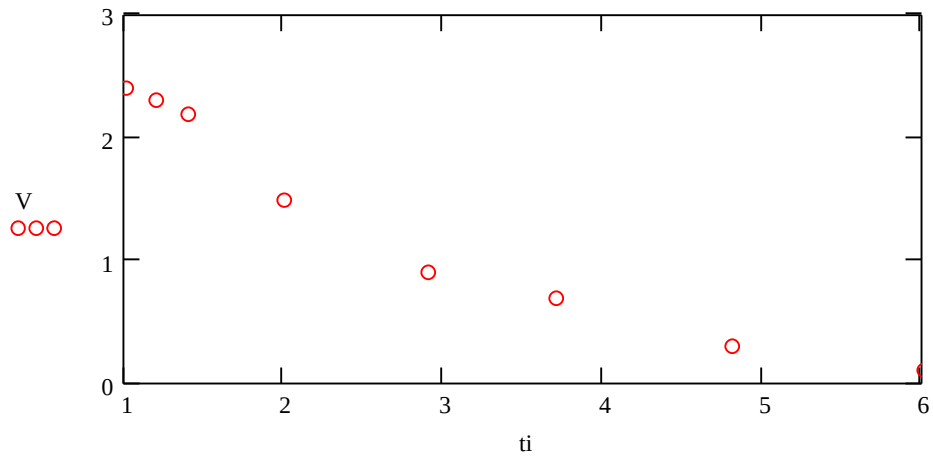
Time, t (s)	Voltage, V (Volts)
1.0	2.4
1.1	2.3
1.4	2.2
2	1.5
2.9	0.9
3.7	0.7
4.8	0.3
6.0	0.1

The following pages are MathCAD worksheets showing the nonlinear transformation and data reduction.

$$ti := \begin{pmatrix} 1 \\ 1.2 \\ 1.4 \\ 2 \\ 2.9 \\ 3.7 \\ 4.8 \\ 6 \end{pmatrix} \qquad V := \begin{pmatrix} 2.4 \\ 2.3 \\ 2.2 \\ 1.5 \\ .9 \\ .7 \\ .3 \\ .1 \end{pmatrix}$$

$n := \mathrm{rows}(ti) \qquad n = 8 \qquad i := 0..n-1$

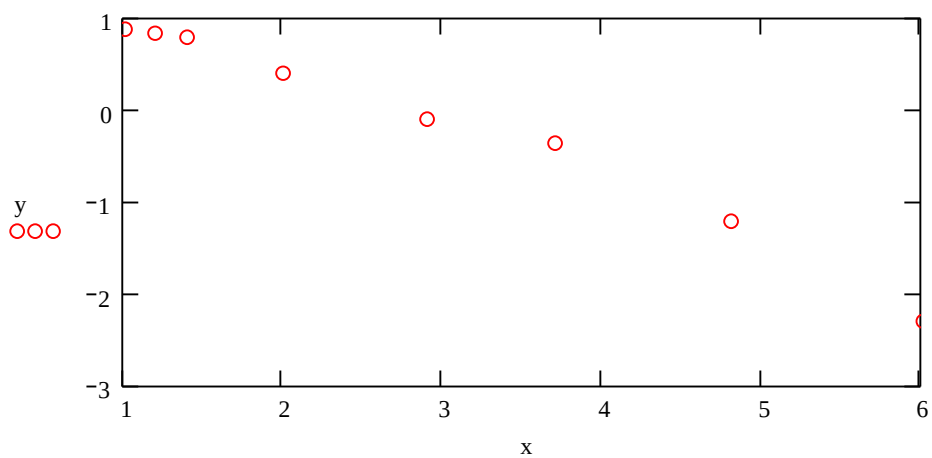
Plot the raw data to see what it looks like



Now do a linearizing transformation

$y := \ln(V) \qquad x := ti$

Plot the transformed data to see what it looks like



At this stage we have transformed the data into a linear equation. We now do a complete linear regression, just like in the previous section on linear regression and line fitting.

Calculate the slope and intercept of the best-fit line

$slp := slope(x, y)$ $\quad slp = -0.6135$

$interc := intercept(x, y)$ $\quad interc = 1.6305$

Calculate the correlation coefficient and standard error

$corr(x, y) = -0.9899$

$stdE := stderr(x, y)$ $\quad stdE = 0.1740$

We compare the correlation coefficent with tabulated values.

Calculated = 0.9899 Tabulated (at 5% level of significance and n=8) = 0.707
Since calculated > tabulated, we conclude that there is no strong evidence that a trend does not exist

Now on to the confidence interval. Look up Student's-t for dof=n-2

$t := qt(.975, n - 2)$ $\quad t = 2.4469$ in the 'qt' function, the first number is 1-(α/2)

$CI := t \cdot stdE$ CI is the confidence interval $\quad CI = 0.4257$

Set up everything for plotting

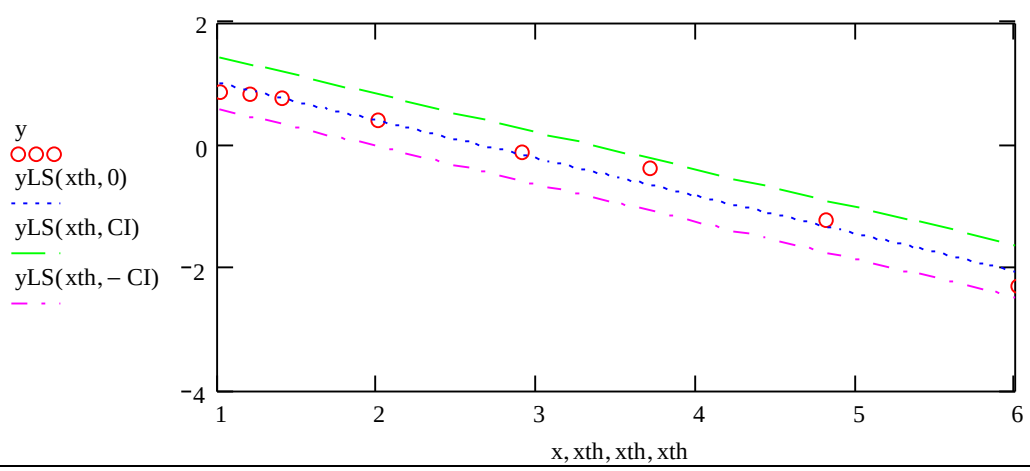
$yLS(xLS, d) := interc + slp \cdot xLS + d$ The straight line equation with 'extra' intercept

In this example, the best fit line and +/- confidence interval lines are plotted by defining their start and end points (2 points define a line!!)

$xth_0 := min(x)$ This is the smallest X value, and is the first x-point for our fitted lines

$xth_1 := max(x)$ This is the biggest (and second) x-point for our least squares lines

Plot the raw data, best-fit line and confidence interval



But we have only found the best-fit line to the *transformed* data:

slope = -0.6135
intercept = 1.6305

We need to "untransform" the slope and intercept so that we can get the values we want. Recall that the linearizing transformation for this example took the form:

Original equation:

$$V = Ge^{-ct}$$

Take natural logs on both sides to linearize:

$$\ln(V) = \ln(G) - ct$$
$$y = b + ax$$

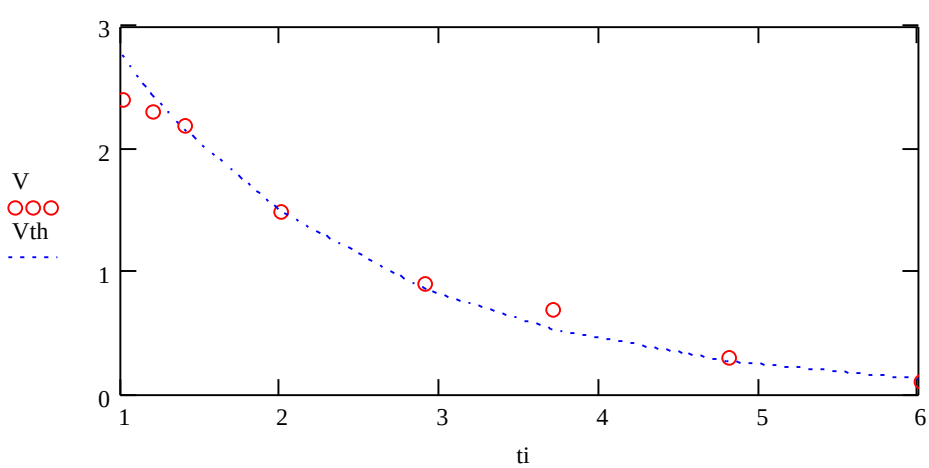
Thus the slope of the best-fit line relates to "-*c*" and the intercept relates to "ln(*G*)". The "untransformation" is:

$$c = -(\text{slope}) = 0.6135$$
$$G = e^{INTERCEPT} = e^{1.6305} = 5.1062$$

Finally, we plot the original raw data and the theoretical curve based on the calculated parameters:

$c := -slp$ $\quad G := e^{interc}$

$Vth := G \cdot e^{-c \cdot ti}$



How do we choose a good linearizing transformation?

Choosing a good linearizing transformation is often straightforward. For example, if you have measured (*N*, *t*) data and want to fit the model:

$$N = ct^d$$

Taking logs of both sides will yield:

$$\log(N) = \log(c) + d.\log(t)$$

The log(N) of your data relates to the transformed y-axis and log(t) is the transformed x-axis. So in this case you would take logs of both your *N* and *t* data. The best-fit line to log(N) vs. log(t) would have:

slope = d
intercept = log(c)

and the "untransformation will yield:

d = slope
c = $10^{\text{intercept}}$ or $e^{\text{intercept}}$ depending upon whether you took base-10 or natural logs

EXAMPLE: As another example, suppose you were testing a material sample and have measured stress, s_n, and stretch ratio, l. You wish to fit the theoretical Mooney formula to the data to see if Mooney's formula is suitable for your material. You know the Mooney formula is:

$$s_n = 2\left[l - \frac{1}{l^2}\right]\left[C_1 + \frac{C_2}{l}\right]$$

What linearizing transformation could you use? Recall, the aim is to use measured s_n and l data to determine the best-fit parameters C_1 and C_2.

The following approach would work. Rewrite the Mooney formula as:

$$\frac{s_n}{2\left[l - \frac{1}{l^2}\right]} = C_1 + \frac{C_2}{l}$$

This is now a linearized equation:

$$\frac{s_n}{2\left[l - \frac{1}{l^2}\right]} = C_1 \quad + C_2 \frac{1}{l}$$

$$y = b \quad + ax$$

Transformed x-data: $x = \frac{1}{l}$

Transformed y-data: $y = \frac{s_n}{2\left[l - \frac{1}{l^2}\right]}$

Slope: C_2

Intercept: C_1

The "untransformation" for this particular linearizing transformation is simple:

C_1 = intercept

C_2 = slope

PROCEDURE FOR DETERMINING A BEST-FIT STRAIGHT LINE USING

NONLINEAR
CORRELATION AND REGRESSION

1. Plot the raw data.
2. Make a subjective decision about the data and any trends.

3. Choose a linearizing transformation. Transform the data to new x- and y- values.

4. Do a COMPLETE linear correlation and regression to the transformed data. From the previous handout this is:

    1. Plot the raw data.
    2. Make a subjective decision about the data and any trends.

    3. Calculate correlation coefficient r_{xy} and compare it to the values in the table. Make a statistical decision.

    4. Calculate slope & intercept of the best-fit line.
    5. Calculate standard error of estimate.
    6. Calculate confidence interval of the intercept.

    7. Plot raw data and overlay the best-fit line with the ±confidence intervals.

        Remember that most data points MUST fall inside the CI
        Expect a few points to fall outside the CI

    INTERMEDIATE LINEAR DECISION BASED ON
        SUBJECTIVE – LOOK LIKE A STRAIGHT LINE?
        STATISTICAL – CORRELATION COEFFICIENT

5. "Untransform" the slope and intercept to the parameters you are looking for.

6. Plot the raw data and the theoretical nonlinear curve based on the identified parameters.

FINAL DECISION BASED ON
    SUBJECTIVE and STATISTICAL assessment of the linearized analysis

    SUBJECTIVE assessment of the raw data and overlayed nonlinear equation.

PROBLEMS

NlinReg-1:
The following data give the voltage measured across a resistor/capacitor circuit.

Time (s)	0.0	0.1	0.2	0.3	0.4	0.5
Voltage (V)	50.0	17.0	7.0	2.0	1.0	0.5

a) Check the theory that the voltage and time are related by the equation

$$V = at + b$$

Do this by completing a full linear regression for the voltage versus time data.

b) Check the theory that the voltage and time are related by the equation

$$V = V_O e^{ct}$$

Do this by completing a full nonlinear regression. Choose a suitable linearizing transformation.

c) What is your conclusion, and why?