10. (Allometric Modeling)

Many modeling problems are nonlinear. One very important class fits a model in the form of an *allometric* or *power law* model. These models have the form

$$y = kx^a. \tag{1}$$

By taking logarithms of this equation and using the properties of logarithms, Eqn. (3.1) can be written in the following form

$$\ln(y) = a\ln(x) + \ln(k).$$

By defining the variables $Y = \ln(y)$, $X = \ln(x)$, and $K = \ln(k)$, then this equation has the form

$$Y = aX + K,$$

which is a straight line with slope a and intercept K. The techniques from before can be used to find the best linear least squares best fit through the logarithm of the data to find the best *allometric model*. This gives an easy method for finding a and K or equivalently k.

Excel's Trendline has the option *Power*, which gives the best fitting power law model using the algorithm above. The user should be aware that the Excel Trendline Power fit is actually finding the linear least squares fit to the logarithms of the data.

The technique of using the linear least squares fit to the logarithms of the data gives the easiest way to fit an allometric model, since virtually all computer applications have linear least squares fitting algorithms. However, this method is finding the least squares best fit to the logarithms of the data and not the least squares best fit directly from the data. The latter method requires a nonlinear least squares best fit of the data. Though the nonlinear least squares best fit is technically a better unbiased fit to the data, it is much more challenging to find. In particular, it usually requires a reasonable initial guess of the parameter values. Below we show how to use Excel to find this nonlinear least squares fit to data.

In the early years the numbers of AIDS cases grew rapidly, and the cumulative cases of AIDS could be modeled by an allometric model of the form

$$c = kt^r,$$

where t is the number of years after 1980, c is cumulative number of AIDS cases, and k and r are parameters that fit the data.

The table below shows the data for the cumulative number of AIDS cases in the US over the time period of 1981-1992. During these early years the disease was not understood, and its spread was very rapid.

Year	Cumulative AIDS Cases	Year	Cumulative AIDS Cases	Year	Cumulative AIDS Cases
1981	97	1985	15,242	1989	120,612
1982	709	1986	29,944	1990	161,711
1983	2,698	1987	52,902	1991	206,247
1984	6,928	1988	83,903	1992	257,085

*Data of the Cumulative AIDS (in thousands of cases)
throughout the US from 1981 until 1992.*

The data clearly cannot be fit well with a linear model, so one option is to use an allometric model. If we define the variable t to be the number of years after 1980 and c to be the cumulative number of AIDS cases, then an appropriate allometric model is given by:

$$c = kt^r,$$

where the parameters k and r are found from the data above. With the use of logarithms this model can be written:

$$\ln(c) = \ln(k) + r\ln(t).$$

If we let $C = \ln(c)$, $K = \ln(k)$, and $T = \ln(t)$, then the above model is written:

$$C = K + rT,$$

which is just a linear equation. Thus, the logarithms of the data can be fit to a straight line, as was done before, to give the allometric model.

Excel for Power Law Fit

The data from our table is entered into an Excel spreadsheet and rearranged to form two columns (A and B) with the first being the Year and the second being the Cumulative AIDS Cases. A column (B) is inserted between these by right clicking on the B in the second column and selecting Insert. This column is labeled t, and its entries are generated by typing '= A2 - 1980' and pulling down to the end of the data. Columns B and C are highlighted, and the user Inserts a Scatter plot. Excel's Trendline is invoked from either the Main menu or by right clicking on a data point. This time in the Format Trendline window, the user checks the Power and Display Equation on chart options. More digits for the formula are obtained by right clicking on the equation and selecting Format Trendline Label... The window that opens allows Number and Scientific to be selected, and Decimal places: 4 is chosen. Appropriate graph titles and axis labels are added and modified, and the equation label can be modified to show the correct variables and look better. Fig. 3.5 on the left shows this graph.
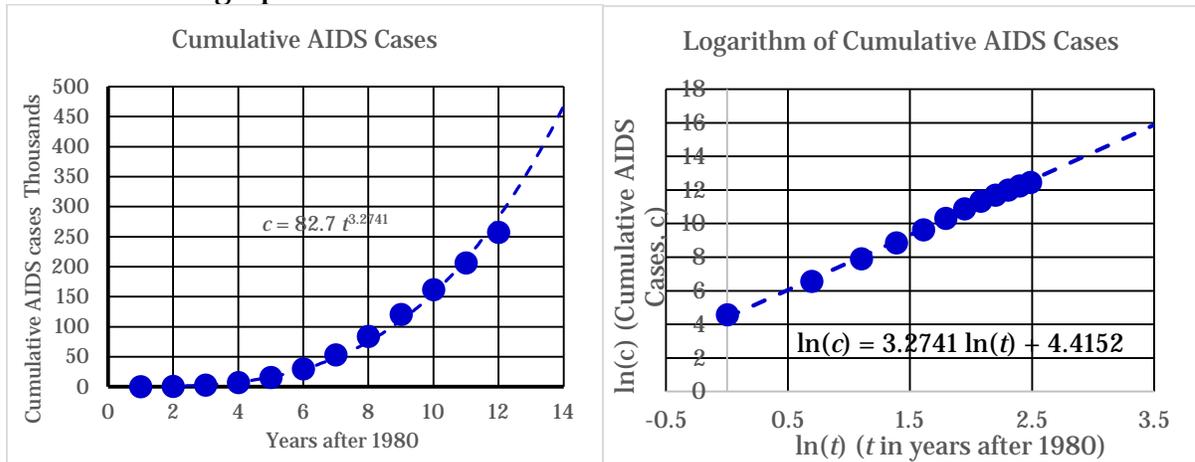
The left graph shows cumulative AIDS cases with the Excel Trendline Power law fit.
The right graph shows the logarithms of the data with the Excel Trendline linear fit.

The model is entered in Column D and pulled down to the end of the data. Column E is used to compute the sum of square errors by entering '= (C2 - D2)^2' and pulling down this formula. Below the last cell, the user clicks on the Σ symbol from either the

Home or Formulas tabs to yield '= SUM(E2:E13)' with this sum of square errors being 932,060,425. This best allometric model is given by:

$$c = 82.700 \ t^{3.2741}.$$

The graph on the right shows the linear least squares best fit to the logarithms of the data. In Columns G and H enter '= ln(B2)' and '= ln(C2),' respectively, and pull down to the end of the data. These columns are highlighted, and the user once again Inserts a Scatter plot. This time Excel's Trendline with a Linear option is applied to give the linear least squares best fit. The graph is improved with the usual techniques. We make a couple of observations about this graph that are relevant to an allometric model. First, it is easy to see that the data appear to lie in a line, so an allometric model is appropriate whenever the graph of the logarithms of the data look almost linear. Second, we observe the formula of the Linear fit and see that the slope of the line (3.2741) agrees with the exponent of the Power law fit. Also, if we exponentiate the value of the intercept from the Linear fit ('= EXP(4.4152)'), then the coefficient of the Power law (82.700) is obtained.

The power law fit usually fails to give the least sum of square errors between the model and the data. To find this Nonlinear Least Squares best fitting model we use Excel's Solver. For the sake of clarity, assume that we are working with the Excel spreadsheet used above for the Cumulative AIDS cases, where Column A has Years 1981-1992, Column B is the years after 1980, Column C has data for the Cumulative AIDS cases, Column D has the allometric model from Trendline's Power law fit, and Column E has the sum of square errors between the data and the Power law model. Insert labels k and r in Cells I2 and I3 and give them the values from the allometric model in Cells J2 and J3. Highlight the four Cells I2:J3 and from the Main menu tab Formulas select Create from Selection: Left column to name the variables k and r. (Note that the variable r becomes $r\_$, because r is reserved for rows.) Label Columns F and G with Nonlin Model and SSE, respectively. In F2, enter '= k*B2^r_ ' and pull down to the end of the data (F13). In G2, enter '= (C2 - F2)^2' and pull down to the end of the data. Below the last cell, click on the Σ symbol from either the Home or Formulas tabs, which gives '= SUM(G2:G13).'

The next step is to invoke Excel's *Solver* to minimize the sum of square errors in G14. From the Main menu under the tab Data, we find *Solver* on the right side with the category Analysis. By clicking on this, a window appears and we Set Objective: to G14 and check the box to Min. Finally, click in the box By Changing Variable Cells: and highlight the Cells J2 and J3. After clicking on the Solve button, the sum of square errors is reduced to 210,542,992, which is less than a quarter the sum of square errors computed by the Power law model, which used the best linear fit to the logarithms of the data. The result gives $k = 213.025$ and $r = 2.8630$, which gives the best allometric model as

$$c = 213.025 \ t^{2.8630}.$$

It is clear that the two techniques give fairly different models.

There are only twelve points in the data set, and graphing the models should include approximately 50 points. Choose three successive blank columns and label them *t*,

Power model, and *Nonlinear model*. Suppose that the label *t* is in Cell L1. In Cell L2, enter 0, then in Cell L3, enter '= L2 + 0.2.' Pull down L3 to L72, where t = 14. In Columns M and N, enter the two versions of the allometric model. Columns L, M, and N are highlighted, and a Scatter plot is created. After the usual improvements are made to the graph, the left graph in Fig. 3.6 is produced. To produce the graph on the right in Fig. 3.6, we copy the graph we produced on the left and Paste Special, using the selection Microsoft Office Drawing Object. Next we either right click on one of the axes or use the Main menu Chart Tools under Layout, selecting either Primary Vertical or Horizontal Axis and choosing the option Show Axis with Log Scale. A warning will appear that *Negative or zero values cannot be plotted correctly on log charts...*, which can be ignored. Since the gridlines are fairly sparse, we have Excel also include the Major & Minor Gridlines. With minor adjustments the log-log graph of Fig. 3.6 is produced. On this graph the data appear to lie on a straight line, which again indicates that an allometric model is a good choice for fitting the data.

In Fig. 3.6, both models fit the data quite well. However, the Trendline Power law fit to the data does a poorer job fitting the larger *t* values. Thus, the Nonlinear Least Squares fit model is better at projecting future trends for cumulative AIDS cases. It is often a subjective decision as to which model is better. In allometric modeling, the most important parameter is the exponent, *r*, and for the two fits in Fig. 3.6 these differ by about 15%, which is not too bad for biological models.

Dog Measurement Study: Allometric models are used to study the relationship between length, weight, and surface area of several dogs. Allometric modeling provides a method of understanding physical relationships between different methods of measurement. This problem examines the relationship of the length of a dog to its weight and surface area. Physiologically, it is important to note that all dogs have a basic shape that one recognizes as a dog. This is a type of self-similarity. A collection of dogs was measured and weighed, producing the table of data below.

a. The first model examines Weight, *w*, as a function of the Length, *u*. Use the allometric modeling technique of fitting the logarithms of the data (Excel's Trendline) to find the best model fitting the data for a model of the form:

$$w = k \, u^a.$$

Length (cm)	Body Weight (gm)	Surface Area (cm2)
48	3460	2245
63	5150	3250
72	5350	3865
76	9980	5000
97	17220	7900
101	26070	8995
105	33260	10535

i) What are the best fit *Trendline* coefficients, *k* and *a* ?
ii) Determine the sum of square errors between the weight model and the data.

iii) If we assume that this model gives the ideal weight for a dog, then how much should a dog with a length of 72 cm weigh?

iv) Assuming this model gives the best value, compute the percent error that the 72 cm dog in the table varies from the ideal dog in weight.

v) Suppose a dog has a length of 97 cm, then how much should this dog weigh?

vi) Again, assuming this model gives the best value, compute the percent error that the 97 cm dog in the table varies from the ideal dog in weight.

vii) Suppose you have a dog that weighs 15 kg. According to the model, what would you predict would be its length?

viii) Create a plot of this best fitting curve with the data. What are the appropriate units for the coefficient k?

b. Repeat the process in Part a for the surface area data, s, as a function of length, u. In this case, the model satisfies the allometric model:

$$s = k\, u^a,$$

where the coefficients k and a differ from Part a.

i) What are the best fit *Trendline* coefficients, k and a ?

ii) Determine the sum of square errors between the surface area model and the data.

iii) If we assume that this model gives the ideal surface area for a dog, then what should be the surface area of a dog with a length of 72 cm?

iv) Assuming this model gives the best value, compute the percent error that the 72 cm dog in the table varies from the ideal dog in surface area.

v) Suppose a dog has a length of 101 cm, then how much surface area should this dog have?

vi) Again, assuming this model gives the best value, compute the percent error that the 97 cm dog in the table varies from the ideal dog in surface area.

vii) Suppose you have a dog that has a surface area of 6400 cm². According to the model, what would you predict would be its length?

viii) Create a plot of this best fitting curve with the data. What are the appropriate units for the coefficient k? Explain why the coefficient a has the value it does when fitting the surface area data for the dogs.

c. Repeat the process in Part a for finding Weight, w, as a function of Length, u. However, use the allometric modeling technique of nonlinear least squares given above to fit the data to the model:

$$s = k\, u^a,$$

i) What are the best fit *nonlinear* fit coefficients, k and a ?

ii) Determine the sum of square errors between the surface area model and the data.

iii) If we assume that this model gives the ideal surface area for a dog, then what should be the surface area of a dog with a length of 72 cm?

iv) Assuming this model gives the best value, compute the percent error that the 72 cm dog in the table varies from the ideal dog in surface area.

v) Suppose a dog has a length of 101 cm, then how much surface area should this dog have?

vi) Again, assuming this model gives the best value, compute the percent error that the 97 cm dog in the table varies from the ideal dog in surface area.

vii) Suppose you have a dog that has a surface area of 6400 cm². According to the model, what would you predict would be its length?

viii) Create a plot of this best fitting curve with the data. What are the appropriate units for the coefficient k? Explain why the coefficient a has the value it does when fitting the surface area data for the dogs.

d. Repeat the process in Part b for finding the surface area, s, as a function of the Length, u. However, use the allometric modeling technique of nonlinear least squares to fit the data to the model:

$$s = k \, u^a,$$

where the coefficients k and a differ from Part b.

i) What are the best fit *nonlinear* fit coefficients, k and a?

ii) Determine the sum of square errors between the surface area model and the data.

iii) If we assume that this model gives the ideal surface area for a dog, then what should be the surface area of a dog with a length of 72 cm?

iv) Assuming this model gives the best value, compute the percent error that the 72 cm dog in the table varies from the ideal dog in surface area.

v) Suppose a dog has a length of 101 cm, then how much surface area should this dog have?

vi) Again, assuming this model gives the best value, compute the percent error that the 97 cm dog in the table varies from the ideal dog in surface area.

vii) Suppose you have a dog that has a surface area of 6400 cm². According to the model, what would you predict would be its length?

viii) Create a plot of this best fitting curve with the data. What are the appropriate units for the coefficient k? Explain why the coefficient a has the value it does when fitting the surface area data for the dogs.