longer the systems stray from stability. Because of this, the models are not as broadly useful as those involving the solution of differential equations, and they typically are used where stable states are obtained relatively rapidly. In later chapters equilibrium models will be combined with the use of differential equations to form other simulation models.

Many of the derivations in this chapter depended on the use of rate equations. With the stability approach, the differentials are eliminated when rates of formation are set equal to rates of breakdown. The use of rate equations is fundamental to the development of many models useful in biological simulation, and examples will be found in subsequent chapters.

CHAPTER 3
ESTIMATING MODEL COEFFICIENTS FROM EXPERIMENTAL DATA

In the two previous chapters you learned how model equations may be derived analytically from theory. You may have noticed that the discussion covered only the form of the equation. When you actually worked with the models in the exercises, you used not only the equations, but also some coefficients that were simply given to you without explanation. In this chapter we describe how to find these coefficients.

At the outset of this discussion of models and data, you should understand clearly the differences between theoretical models and empirical models. The equation for a theoretical model is obtained from ideas about how some biological process works. These ideas may be based only on a loose analogy to known processes, but they still lead to an analytical derivation of an equation. The models in the first two chapters were derived from theory.

Empirical equations, on the other hand, do not have a theoretical origin. They are used when the modeler is not interested in the cause-effect relationship of variables used in the equation, but wishes to use the output as part of a larger model. For example, suppose you are making a rather complicated model of plankton populations in a lake, and you need an equation that describes the density of water at different temperatures. There are, of course, many theoretical models for the effect of temperature on water density, but in constructing the plankton model it may not be essential to know this theoretical basis. In such a case, a usable equation and coefficients may be obtained directly from experimental data. The most commonly used empirical equations are of the polynomial type, and their method of derivation is discussed below. Like theoretical models, empirical models require accurate coefficients if they are to be useful.

From a mathematical viewpoint, an empirical equation and a theory-based equation are equally useful for modeling a particular system, assuming they are equally accurate. However, most modelers would prefer to
use equations that have some basis in theory, because this will contribute more to an understanding how the system functions.

In general, coefficients for both theoretical and empirical models are obtained by the process of "curve fitting". In the case of theoretical models, we first formulate the model from our conception of how the system might work, next we plot data from the system, and then we try to find coefficients for our equation so that the line drawn by our equation best "fits" the actual data. With empirical equations, we plot the data and then look simultaneously for an equation and coefficients that produce a best-fitting line. Notice the difference in sequence. Thus the sequence for theoretical models is equation, then data, then coefficients, while for empirical models it is data, then equation and coefficients. Because these differences are fundamental, this chapter will be divided into two parts, the first considering theoretical models (Sections 3.1 through 3.6), and the second empirical models (Sections 3.7 and 3.8).

3.1 Selecting a Model Equation

In practice, the sequence for developing theoretical models is sometimes scrambled. We may not have available the knowledge or insight required to formulate an analytical model directly. Unlike our model for cell growth, built in Chapter 1 from ideas about how cells grow, in some cases we may need to collect data to get an idea about how the system works. There is no real "Start Here" point in the overall research process of Figure I.1. Thus, data may suggest an equation which can give us a conceptual notion of the workings of the process.

A first step in working with a theoretical model, even if you are not sure of its final mathematical form, is to plot the data to produce a scatter plot, also referred to as an approximating curve. The shape of this curve will indicate whether you developed your initial model correctly, or it may suggest some equations that describe the phenomenon of interest. Some characteristics of the approximating curve will help in selecting an equation or determining whether an analytical equation is likely to be correct. These are: linearity, presence of minima or maxima, direction of slope, and presence or absence of inflection points. A very large number of equations can be formulated to describe any particular line. In practice, we usually select an equation from several simple equations that have proven useful in biological explanation in other cases that are more or less similar.

For example, suppose we wished to model the increase of biomass in a cell culture based on the following hypothetical data:

<table>
<thead>
<tr>
<th>Time</th>
<th>Biomass</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9.9</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>89</td>
</tr>
<tr>
<td>3</td>
<td>271</td>
</tr>
<tr>
<td>4</td>
<td>808</td>
</tr>
</tbody>
</table>

When these data are plotted, they appear to follow an exponential growth curve. As you know from Chapter 1, the curve is described by the equation

\[ y = Ae^{Bx} \]  

with \( x \) representing elapsed time, \( A \) the biomass at \( x = 0 \), and \( B \) the rate of growth. The general problem of this chapter is to find the best values for the coefficient \( A \) and the exponent \( B \). The two prevalent approaches to this problem are the use of linear transformation, and nonlinear regression. The techniques for these procedures are discussed in the following sections.

3.2 Linear Transformation Procedures

The technique of linear transformation takes advantage of the fact that many useful equations may be rearranged to a linear form. Experimental data that conform to the equation may then be transformed and fitted to straight line of the transformed equation by the usual linear least-squares methods. The slope and intercept of the resulting straight line equation may be used to find estimates of the values of the coefficients and exponents of the original equations.

In the example case above, the exponential growth equation may be converted to linear form by taking the natural log of both sides to give

\[ \ln(y) = \ln(A) + Bx \]  

Now, if values of \( y \) from the data are transformed to \( \ln(y) \) and are plotted against time \( x \), the points should fall on a straight line that has a slope of \( B \) and an intercept of \( \ln(A) \). (You will recall from elementary algebra the expression for a straight line, \( y = i + sx \), where \( i \) and \( s \) are the intercept and slope.) This transformation from the exponential is frequently used to find the growth rate of microorganisms. In this example, the transformed data become
When the log(Biomass) data are plotted against time we obtain the straight line shown in Figure 3.1.

\[
\begin{array}{ccc}
\text{Time} & \text{Biomass} & \text{log(Biomass)} \\
0 & 9.9 & 2.3 \\
1 & 32 & 3.4 \\
2 & 89 & 4.5 \\
3 & 271 & 5.6 \\
4 & 808 & 6.7 \\
\end{array}
\]

The differences, also called residuals or deviations, appear to be relatively small and randomly distributed; the exponential equation with the fitted values of \(A\) and \(B\) appears to be a reasonable predictor of growth for the organisms. In linear regression, the residual values are important in assessing the fit of the equation to the data.

### 3.3 Notes on Regression with Linear Transformation

Linear regression should be familiar to all biologists, and is discussed thoroughly in most elementary texts of biological statistics (e.g. Zar 1984). The line obtained with the linear regression procedure minimizes the sum of the squares of the vertical distances (residuals) between the data points and the regression line. In addition to values for slope and intercept, several other statistics based on the residuals are usually obtained from a regression analysis to determine how well the line fits the data.

In the usual linear regression procedure, the fit of the line is assessed from the residuals. While this is also the case for transformed data, the residuals are calculated after back-transformation, using the nonlinear form of the equation. Because you will use the procedure in several exercises in this chapter, it will be mentioned here briefly. Consult your favorite biometry textbook for further discussion.

The residual sum of squares is found from the differences between the observed and predicted values of \(y\), and is given by

\[
SS_{\text{residual}} = \sum (y_i - \hat{y}_i)^2
\]  

with the predicted and observed values derived from the actual data and the theoretical equation. The variance, or residual mean square, is given with

\[
MS_{\text{residual}} = \frac{SS_{\text{residual}}}{n - P}
\]  

where \(n\) is the number of \(x - y\) pairs and \(P\) is the number of parameters (coefficients and exponents) obtained by the fitting procedure. In most cases of linear transformation, \(P\) has the value of two, for \(A\) and \(B\). The standard error of the regression estimate is the square root of the residual mean square.

An \(F\)-statistic, used with an \(F\)-table to assess whether equations explain significant amounts of variation between the \(y\)-values and the overall mean \(y\), is found with

\[
F = \frac{SS_{\text{regression}}}{SS_{\text{residual}}} \cdot \frac{n - P}{P - 1}
\]

Note again that these statistics are found using the residuals from values predicted with the original, untransformed equation. They will differ from
statistics calculated from the residuals of the transformed data and the linear equation.

### 3.4 Useful Equations for Linear Transformation

A variety of equations that produce a diversity of curve shapes may be linearized with transformation procedures; see Hoerl (1954) for examples. A number of these equations that have been useful in biological research are given in Table 3.1. Each equation is presented in both the nonlinear standard form, and in the linear form.

Some forms of the curves produced by these equations are shown in Figure 3.2.

The values of the parameters producing the curves are given to show the role they play in determining the shape of the curve.

You will notice that the same general shape of curve may be produced by equations with different forms. For example, a negatively accelerated rise to a plateau may be produced by the hyperbolic (Eq. 2), reciprocal exponential (Eq. 5), and exponential saturation (Eq. 7) equations. Thus, before undertaking a curve-fitting analysis you should have a good idea about the mechanism that may produce your results.

Several commercial programs for large and small computers are available that can transform sets of x-y data and perform a linear regression for several of the equations in Table 3.1. We have incorporated these curvilinear equations, along with the straight-line equation into a BASIC program, CURFIT. The program can assess the fit of sets of x-y data to these equations, and determine the best-fit coefficients. The procedures used in running this program are described in Appendix 3. The program is designed to make it easy to perform the different transformations of data, and to assess the goodness of fit of the back-transformed data, as described above.

If the specific equation is not known from theoretical equations, the CURFIT program permits a quick comparison among several that might describe the set of data. (If an equation is selected in this manner, you should not attach a great deal of weight to the biological mechanism implied by the equation.) The shape of the curves in Figure 3.2 may help you to decide which equations to select in fitting data with the CURFIT program.

Different statistics are used to compare and assess the relative goodness of fit among several equations. It is generally held that the best fitting equation will have the minimum residual mean square. This statistic is better than the residual sum of squares because it takes into account the different degrees of freedom associated with different model equations. Other statistics which may be employed to evaluate goodness of fit include the F-statistic and the standard error of the regression estimate.

The program finds the coefficient of determination, $R^2$, for the linear equation. This value, which has a maximum of 1, compares variance of the predicted y-values with the variance of the actual y-values. This coefficient is not a direct function of fit when nonlinear equations are involved. In fact, for nonlinear equations, higher $R^2$ values may be obtained from equations that are obviously poorly fit. Also, values of $R^2$ that exceed 1.0 may be obtained when the variance of data predicted by the nonlinear model actually exceeds that of the original data.

### 8.5 Notes on Some of the Equations

Several of the equations listed in Table 3.1 and Figure 3.2 and used in the CURFIT program deserve some comment. The hyperbolic equation, for example, should be familiar from our discussion of Michaelis-Menten kinetics in Chapter 2. The equation may be linearized in a variety of ways. The usual Lineweaver-Burk plot (see Figure 3.3) is the poorest of the possibilities (Dowd and Riggs 1965). It is based on the following equation:

$$\frac{1}{y} = \frac{1}{A} + \frac{(B/A)}{x}$$

(3.6)

The Woolf transformation (Table 3.1) used in CURFIT is a better estimator of the parameter values. Raaijmakers (1987) provides a recent review of the problems in estimating enzyme kinetics with the equations.

Several of the curves used in CURFIT require an estimate for a constant used in the equation. These are usually an asymptote or an intercept of the approximating curve. A trial value to enter to begin the analysis can be obtained by a visual approximation of the value from a rough curve. From the initial value, the curve-fitting process may be repeated to find the value for the constant that maximizes the value of the $F$-statistic or produces the minimal residual mean square.

The exponential saturation equation (Eq. 7 in Table 3.1) appeared in Chapter 1 as the Von Bertalanffy growth model. The equation requires an estimated value for $A$ in order to calculate the slope $(\ln A)$ and intercept of the linear form of the equation. Note that $A$-value appears twice in the equation. If you select this equation to fit to your data, you may have to repeat the curve-fitting procedure several times so that your estimate of $A$ equals the computed value.

Equation 10, labelled “sigmoid”, can in fact produce a variety of curve shapes other than sigmoidal. This is shown in Figure 3.2. This equation can have rather high values of $F$ for a variety of data sets. (This provides another argument for knowing which equation is likely to explain the underlying mechanism of your system.)
<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Standard Function</th>
<th>Linear Transformation</th>
<th>Definition of Parameters by Slope (S) and Intercept (I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>straight line</td>
<td>$y = A + Bx$</td>
<td>$[y] = A + B[x]$</td>
<td>S I</td>
</tr>
<tr>
<td>2.</td>
<td>hyperbolic</td>
<td>$y = A x / (B + x)$</td>
<td>$[x/y] = (B/A) + (1/A)[x]$</td>
<td>1/S I/S</td>
</tr>
<tr>
<td>3.</td>
<td>modified inverse</td>
<td>$y = A / (B + x)$</td>
<td>$[1/y] = (B/A) + (1/A)[x]$</td>
<td>1/S I/S</td>
</tr>
<tr>
<td>4.</td>
<td>exponential</td>
<td>$y = A \cdot \exp(Bx)$</td>
<td>$[\log_e y] = \log_e A + B[x]$</td>
<td>exp(I) S</td>
</tr>
<tr>
<td>5.</td>
<td>exp. reciprocal</td>
<td>$y = A \cdot \exp(B/x)$</td>
<td>$[\log_e y] = \log_e A + B[1/x]$</td>
<td>exp(I) S</td>
</tr>
<tr>
<td>6.</td>
<td>maxima function</td>
<td>$y = A x \cdot \exp(Bx)$</td>
<td>$[\log_e (y/x)] = \log_e A + B[x]$</td>
<td>exp(I) S</td>
</tr>
<tr>
<td>7.</td>
<td>exp. saturation</td>
<td>$y = A [1 - \exp(Bx)]$</td>
<td>$[\log_e (A - y)] = \log_e A + B[x]$</td>
<td>exp(I) S</td>
</tr>
<tr>
<td>8.</td>
<td>logistic</td>
<td>$y = K / [1 + A \cdot \exp(Bx)]$</td>
<td>$[\log_e (K/y - 1)] = \log_e A + B[x]$</td>
<td>exp(I) S estimate</td>
</tr>
<tr>
<td>9.</td>
<td>logarithmic</td>
<td>$y = K + (A x^B)$</td>
<td>$[\log_e y] = \log_e A + B[\log_e x]$</td>
<td>exp(I) S estimate</td>
</tr>
<tr>
<td>10.</td>
<td>“sigmoid”</td>
<td>$y = K / (1 + A x^B)$</td>
<td>$[\log_e (K/y - 1)] = \log_e A + B[\log_e x]$</td>
<td>exp(I) S estimate</td>
</tr>
</tbody>
</table>

Table 3.1. Listing of some equations that occur often in biological and biomedical research, and their linearized forms. The transformed values of $x$ and $y$ are given in brackets in the linear form. The definitions of parameters show how they are derived from the linear form. Equation numbers refer to the CURFIT program.
Figure 3.2. Curves from equations of Table 3.1.
3.6 Practicalities and Problems of Regression with Transformation

Linear regression with transformed data has several attractive features. The procedures are easy to understand, they have a long history in the biological literature, and they can be carried out with pencil and paper with smaller data sets. However, there are some more or less severe statistical difficulties associated with the procedure.

Linear regression is properly performed when the variability in values of y are not related to x or to y. Unfortunately, this is often not the case in biological research, and a transformation may aggravate the problem. In addition, after transformation the data points may have different weights than before transformation.

![Graph](image)

Figure 3.3. Plot of reciprocal of hypothetical data, with uniform spacing of intervals of x. The reciprocal plot results in a non-uniform clustering of data points.

For example, in Figure 3.3, uniformly distributed values of x have been transformed to 1/x, so that most are clustered at low values of 1/x. The points with high 1/x values have a great influence on the slope of the line, which is almost certain to pass through the extreme point. In the hyperbolic curve, this point has the lowest value of both x and y. This point probably has the greatest percentage error, because it is located on the steepest part of the hyperbolic curve.

Other problems arise from the appearance of negative or zero values in the data. Log transformation of such values is impossible, of course,

Estimating Coefficients from Data

As is reciprocal transformation of zero values. Approximating zero with a very small number (e.g. $10^{-6}$) will distort greatly the impact of the value on the linear regression. Such data may be normalized to positive real numbers by adding a constant to all the affected values before linear analysis. Depending on the equation, this may introduce further bias into the estimates of the coefficients. Alternatively, the data pair may simply be dropped from the analysis.

Problems with the analysis of linearly transformed data may be discovered sometimes by inspecting a plot of the residuals vs. the x-value. (URHFTT provides such a plot.) If the residuals are not distributed uniformly, nonlinearity is indicated. If the residual plot tends to look wedge-shaped, it may indicate the overly strong impact of an extreme point. If the residual plot is U-shaped, it may indicate that the transformed data are not well-described by the particular equation.

Exercises: The following sets of data have been synthesized from standard equations. Fit at least two theoretical equations to each using linear transformation techniques. Decide on the best-fitting equation based on either the residual mean square or the F-statistic.

<table>
<thead>
<tr>
<th>Exercise 3.1</th>
<th>Exercise 3.2</th>
<th>Exercise 3.3</th>
<th>Exercise 3.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>33</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>15</td>
<td>43</td>
<td>15</td>
<td>69</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>25</td>
<td>56</td>
<td>25</td>
<td>86</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>30</td>
<td>90</td>
</tr>
<tr>
<td>35</td>
<td>64</td>
<td>35</td>
<td>92</td>
</tr>
<tr>
<td>40</td>
<td>67</td>
<td>40</td>
<td>94</td>
</tr>
<tr>
<td>45</td>
<td>69</td>
<td>45</td>
<td>95</td>
</tr>
<tr>
<td>50</td>
<td>71</td>
<td>50</td>
<td>96</td>
</tr>
<tr>
<td>50</td>
<td>17</td>
<td>50</td>
<td>10.7</td>
</tr>
</tbody>
</table>

3.7 Nonlinear Regression

The use of linear transformation and regression is limited by the problems outlined above. In addition, many useful equations cannot be made linear with a transformation. Nonlinear regression is a useful alternative...
to linear transformation, and may provide the only method for working with some equations. The subject will be given a brief introduction here, to permit some elementary work with CURNLFIT, a nonlinear regression program in Appendix 5.

The objective of nonlinear regression is the same as linear regression: to select values of coefficients that minimize the sum of the squared values of residuals. Thus, nonlinear regression is also a least-squares procedure. However, unlike linear and polynomial regression methods, nonlinear regression requires the computer to make several steps to solve for least-squares coefficients. These steps or iterations begin with a first guess about the values of the coefficients. These values are then adjusted to improve the fit of the curve to the data, and then adjusted again. These iterations will proceed until little improvement is noticed in the values. Several methods of iteration to a solution are commonly available in different computer programs for nonlinear regression; Motulsky and Ransnas (1987) provide a clear and helpful discussion of these from a biologist's perspective.

In using any nonlinear regression program, several decisions about the type of data have to be made, because different calculating procedures are used based on the decisions. Obviously, the initial decision will involve the selection of one or perhaps two equations that might describe the relationship between $x$ and $y$. The data should be scaled so that the units of the data are neither very large nor very small. The regression program will work more accurately with values of 0.1 to 10 micrograms, rather than 0.000001 grams. The user will have to specify accurate estimates for initial values. A poor selection of values will slow the computer's search for the best estimates, and may send the nonlinear regression program on a hunt for values in the wrong direction, so that it never converges, or perhaps converges on incorrect values. In some cases, reasonable estimates of values can be made by inspecting the data. If possible, values from a regression based on linear transformation (e.g. CURFIT) can be used for initial estimates.

The weighting scheme used with the data will be important. Linear regression assumes that the variability in $y$ is unrelated to the magnitude of $x$ and $y$. However, in biological research it is often the case that variability increases with increasing $y$; in many biological procedures, the experimental uncertainty associated with a measured value is related to the size of the value. In this case, data points with large values of $y$ are scattered about the regression line more than points with small $y$ values. In estimating regression coefficients, a least-squares program would emphasize points with larger values of $y$. To avoid this problem with nonlinear regression, different weights may be given very high and very low values of $y$, depending upon the nature of the data. The CURNLFIT

program offers the option of selecting for an assumption of no weighting, or weighting that assumes a variation proportional to values of $y$, or an "in between" weighting scheme. If the variability of measurement is assumed to be proportional to $y$, the program will minimize the sum of squares of the relative distances of the points from the curve, rather than the measured distances. In general, it is probably best to select "no weighting" if you do not have information about the variability of the data.

The problem of how to treat "outlying values" occurs in both linear and nonlinear regression. The point(s) furthest from the line will be the most important in increasing the value of the residual sum of squares. It is tempting simply to discard such values as being mistakes in data collection, and thus to increase the goodness of fit considerably. The problem then is to determine what is a mistake, and what is real variation in the system. A technique that is less biased is to modify the residuals so that less weight is given to outlying values. A variety of such systems have been used. CURNLFIT allows the selection of a bisquare weighting system for outlying values (Duggleby 1981).

In addition to having the equations from CURFIT built into the program, CURNLFIT also permits the user to write into the program any two- or three-parameter nonlinear equation for a least-squares estimate. The details of the procedure are in Appendix 5.

In assessing goodness of fit with nonlinear regression, the plot of residuals is particularly important as a way of comparing more than one equation. Values of $R^2$, the coefficient of determination, are not particularly useful, nor are values of $F$. An $F$-test can be used to compare the fit of two equations; Motulsky and Ransnas (1987) describe the procedure.

Exercises: Use the data and results from the linear regression Exercises 3-1 to 3-4 above to explore the procedures of nonlinear regression. For any of the data sets, pick the one or two best-fitting equations and perform a nonlinear regression using the same equation, and the linear estimates of coefficients as initial values for the iterations.

Exercise 3-5: Use data and results from Exercise 3-1.
Exercise 3-6: Use data and results from Exercise 3-2.
Exercise 3-7: Use data and results from Exercise 3-3.
Exercise 3-8: Use data and results from Exercise 3-4.

3.8 Polynomial Regression

Fitting empirical equations to experimental data is generally easier to
accomplish than fitting theoretical equations because there are fewer decisions to be made. However, from the viewpoint of discovering or explaining the function of a system, polynomials and other empirical equations have little value; they should be used sparingly. If you immediately jump to the use of an empirical equation without trying to use one based on insight and theory, you may be missing up an opportunity to understand some components of the system.

Assuming you are satisfied that an empirical equation is needed, you may find the best fit to the following equation:

\[ y = A_0 + A_1 x + A_2 x^2 + A_3 x^3 + A_4 x^4 + \ldots \] (3.7)

The aim of polynomial regression is to find the values for \( A_0, A_1, \) etc., that make the best fitting curve. When using polynomial regression you must specify the order of the regression, which amounts to the number of coefficients to be fit. When only \( A_0 \) and \( A_1 \) are used, the equation will define a straight line or a first-order equation. The remaining equations formed by adding successive coefficients are quadratic (second-order), cubic (third-order), quartic (fourth-order), and quintic (fifth-order). Ideally, you would fit these equations with the usual procedure of polynomial regression analysis, starting with the lowest order and stopping when the addition of more terms does not improve the fit.

In Appendix 4 of this text you are provided with a curve-fitting program called POLYFIT, which allows you to fit up to five different polynomial equations to a set of \( x-y \) data.

**Exercises:** The following data sets were obtained from various sources in the scientific literature. For each exercise, select at least two theoretical equations from Table 3.1 that appear to match the form of the curve. Then use the CURFIT program or a similar program to analyze the data. Based on your CURFIT results, fit the best equation with nonlinear regression using CURNLFIT or other nonlinear regression program. For those data sets which do not appear to fit any available equation, use the POLYFIT program.

**Exercise 3-9:** Data from Machado et al. (1989)

X: Minutes after injection of radiolabeled glucose

Y: Specific radioactivity of blood glucose in a Brazilian carnivorous fish, *Hoplias malabaricus*, starved for 10 months

**Exercise 3-10:** Data from Sirko et al (1989)

X: minutes after injection of endotoxin

**Exercise 3-11:** Data from Kooloos and Zweers (1989)

X: depth of water in drinking vessel, mm

Y: ml of water swallowed in a single drinking cycle (bill dipped and raised) by mallard ducks (*Anas platyrhynchos*)

**Exercise 3-12:** Data from Fargo and Bonjour (1988)

X: Constant incubation temperature in °C

Y: Developmental rate (time\(^{-1}\)) of third instar nymphs of a squash bug, *Anasa tristis*

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>60.0</td>
<td>0</td>
<td>38.4</td>
<td>2.5</td>
<td>0.7</td>
<td>20.0</td>
<td>0.0885</td>
</tr>
<tr>
<td>15</td>
<td>52.1</td>
<td>30</td>
<td>38.5</td>
<td>5.0</td>
<td>1.6</td>
<td>21.7</td>
<td>0.1162</td>
</tr>
<tr>
<td>30</td>
<td>49.0</td>
<td>60</td>
<td>38.7</td>
<td>10.0</td>
<td>2.0</td>
<td>23.3</td>
<td>0.1724</td>
</tr>
<tr>
<td>45</td>
<td>45.7</td>
<td>90</td>
<td>39.4</td>
<td>12.5</td>
<td>2.3</td>
<td>26.7</td>
<td>0.2778</td>
</tr>
<tr>
<td>60</td>
<td>39.3</td>
<td>120</td>
<td>39.8</td>
<td>17.5</td>
<td>2.5</td>
<td>28.9</td>
<td>0.3333</td>
</tr>
<tr>
<td>90</td>
<td>35.3</td>
<td>150</td>
<td>40.0</td>
<td>22.5</td>
<td>2.7</td>
<td>31.1</td>
<td>0.3571</td>
</tr>
<tr>
<td>120</td>
<td>29.2</td>
<td>180</td>
<td>40.2</td>
<td>27.5</td>
<td>2.8</td>
<td>33.3</td>
<td>0.4167</td>
</tr>
<tr>
<td>180</td>
<td>22.5</td>
<td>210</td>
<td>40.4</td>
<td>30.0</td>
<td>3.2</td>
<td>35.6</td>
<td>0.4348</td>
</tr>
<tr>
<td>240</td>
<td>18.3</td>
<td>240</td>
<td>40.4</td>
<td>35.0</td>
<td>3.1</td>
<td>36.7</td>
<td>0.4000</td>
</tr>
<tr>
<td>300</td>
<td>15.4</td>
<td>40.0</td>
<td>3.2</td>
<td>37.8</td>
<td>0.4000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>360</td>
<td>12.0</td>
<td>38.9</td>
<td>0.4167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Exercise 3-13:** Data from Mukhtar et al. (1989)

X: Concentration (µM) of LTB₄ (leukotriene B₄), a pro-inflammatory agent

Y: Activity of rat hepatic microsomal LTB₄ ω-hydroxylase (pmol min⁻¹ mg⁻¹ protein), measured after 20 min of aerobic incubation of microsomes at 37°C.
Exercise 3-14: Data from Miyazono and Heldin (1989)

X: Log concentration (nM) of transforming growth factor-β1 (TGF-β1) in normal rat kidney cell cultures

Y: Percentage of radiolabeled TGF-β1 remaining bound to cells after incubation with unlabeled TGF-β1.

Exercise 3-15: Data from Tulasi and Ramana Rao (1989)

X: Environmental oxygen tension in mm Hg

Y: Equilibrium percentage oxygen saturation of blood of a freshwater crab, Barytelphusa guerini

Exercise 3-16: Data from Inagaki and Yamashita (1989)

X: Age (days) of fifth-instar larvae of silkworm (Bombyx mori)

Y: Radiolabeled acetate uptake by lipids of larval fat body during 1 hr incubation after surgical removal of fat body of female silkworm larvae

Exercise 3-17: Data from Fukuhara and Takao (1988)

X: Days after hatching

Y: Time (seconds) spent swimming during a 1-minute period by larval anchovy, Engraulis japonica

Exercise 3-17: Data from Lynch (1989)

X: Concentration (μg ml⁻¹) of food in culture water

Y: Mean lifespan (days) of a small freshwater crustacean, Daphnia pulex

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
<td>0.04</td>
<td>14.9</td>
</tr>
<tr>
<td>1</td>
<td>1.5</td>
<td>0.12</td>
<td>19.8</td>
</tr>
<tr>
<td>2</td>
<td>4.1</td>
<td>0.16</td>
<td>36.4</td>
</tr>
<tr>
<td>3</td>
<td>15.9</td>
<td>0.31</td>
<td>32.4</td>
</tr>
<tr>
<td>4</td>
<td>49.5</td>
<td>0.55</td>
<td>51.3</td>
</tr>
<tr>
<td>5</td>
<td>58.5</td>
<td>0.78</td>
<td>53.9</td>
</tr>
<tr>
<td>6</td>
<td>59.5</td>
<td>1.55</td>
<td>44.4</td>
</tr>
</tbody>
</table>

Conclusion

This chapter has shown how to find coefficients of the best fitting theoretical or empirical equations using experimental data. In each case we have assumed that the cause and effect relationship between the dependent (y) variable and the independent (x) variable was not in question. Regression techniques are used here to find coefficients, not to assess the significance of the relationship. The resulting best-fit equation is used only for predicting values of y for a given x. In several exercises throughout this book, you will use the techniques of curve-fitting to find coefficients from data.

The statistics such as F-values that are used by the curve-fitting programs should be used only for establishing the condition of best fit. You should consult any good reference on statistics before you test hypotheses about the relationship of one variable to another using regression.