3 Model Fitting

Introduction

In the mathematical modeling process we encounter situations that cause us to analyze data for different purposes. We have already seen how our assumptions can lead to a model of a particular type. For example, in Chapter 2 when we analyzed the distance required to bring a car to a safe stop once the brakes are applied, our assumptions led to a submodel of the form

$$d_b = Cv^2$$

where d_b is the distance required to stop the car, v is the velocity of the car at the time the brakes are applied, and C is some arbitrary constant of proportionality. At this point we can collect and analyze sufficient data to determine whether the assumptions are reasonable. If they are, we want to determine the constant C that selects the particular member from the family $y = Cv^2$ corresponding to the braking distance submodel.

We may encounter situations in which there are different assumptions leading to different submodels. For example, when studying the motion of a projectile through a medium such as air, we can make different assumptions about the nature of a drag force, such as the drag force being proportional to v or v^2. We might even choose to neglect the drag force completely. As another example, when we are determining how fuel consumption varies with automobile velocity, our different assumptions about the drag force can lead to models that predict that mileage varies as C_1v^{-1} or as C_2v^{-2}. The resulting problem can be thought of in the following way: First, use some collected data to choose C_1 and C_2 in a way that selects the curve from each family that best fits the data, and then choose whichever resultant model is more appropriate for the particular situation under investigation.

A different case arises when the problem is so complex as to prevent the formulation of a model explaining the situation. For instance, if the submodels involve partial differential equations that are not solvable in closed form, there is little hope for constructing a master model that can be solved and analyzed without the aid of a computer. Or there may be so many significant variables involved that one would not even attempt to construct an explicative model. In such cases, experiments may have to be conducted to investigate the behavior of the independent variable(s) within the range of the data points.

The preceding discussion identifies three possible tasks when we are analyzing a collection of data points:

1. Fitting a selected model type or types to the data.
2. Choosing the most appropriate model from competing types that have been fitted. For example, we may need to determine whether the best-fitting exponential model is a better model than the best-fitting polynomial model.
3. Making predictions from the collected data.

In the first two tasks a model or competing models exist that seem to *explain* the observed behavior. We address these two cases in this chapter under the general heading of *model fitting*. However, in the third case, a model does not exist to explain the observed behavior. Rather, there exists a collection of data points that can be used to *predict* the behavior within some range of interest. In essence, we wish to construct an *empirical model* based on the collected data. In Chapter 4 we study such empirical model construction under the general heading of *interpolation*. It is important to understand both the philosophical and the mathematical distinctions between model fitting and interpolation.

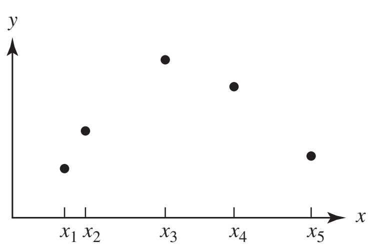
Relationship Between Model Fitting and Interpolation

Let's analyze the three tasks identified in the preceding paragraph to determine what must be done in each case. In Task 1 the precise meaning of *best* model must be identified and the resulting mathematical problem resolved. In Task 2 a criterion is needed for comparing models of different types. In Task 3 a criterion must be established for determining how to make predictions in between the observed data points.

Note the difference in the modeler's attitude in each of these situations. In the two model-fitting tasks a relationship of a particular type is strongly suspected, and the modeler is willing to accept some deviation between the model and the collected data points to have a model that satisfactorily *explains* the situation under investigation. In fact, the modeler expects errors to be present in both the model and the data. On the other hand, when interpolating, the modeler is strongly guided by the data that have been carefully collected and analyzed, and a curve is sought that captures the trend of the data to *predict* in between the data points. Thus, the modeler generally attaches little explicative significance to the interpolating curves. In all situations the modeler may ultimately want to make predictions from the model. However, the modeler tends to emphasize the proposed *models* over the data when model fitting, whereas when interpolating, he or she places greater confidence in the *collected data* and attaches less significance to the form of the model. In a sense, explicative models are *theory* driven, whereas predictive models are *data* driven.

Let's illustrate the preceding ideas with an example. Suppose we are attempting to relate two variables, y and x, and have gathered the data plotted in Figure 3.1. If the

Figure 3.1
Observations relating the variables y and x



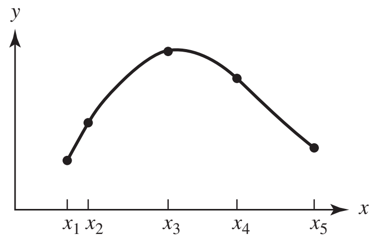


Figure 3.2
Interpolating the data using a smooth polynomial

modeler is going to make predictions based solely on the data in the figure, he or she might use a technique such as *spline interpolation* (which we study in Chapter 4) to pass a smooth polynomial through the points. Note that in Figure 3.2 the interpolating curve passes through the data points and captures the trend of the behavior over the range of observations.

Suppose that in studying the particular behavior depicted in Figure 3.1, the modeler makes assumptions leading to the expectation of a quadratic model, or parabola, of the form $y = C_1x^2 + C_2x + C_3$. In this case the data of Figure 3.1 would be used to determine the arbitrary constants C_1, C_2, and C_3 to select the best parabola (Figure 3.3). The fact that the parabola may deviate from some or all of the data points would be of no concern. Note the difference in the values of the predictions made by the curves in Figures 3.2 and 3.3 in the vicinity of the values x_1 and x_5.

A modeler may find it necessary to fit a model and also to interpolate in the same problem. Using the best-fitting model of a given type may prove unwieldy, or even impossible, for subsequent analysis involving operations such as integration or differentiation. In such situations the model may be replaced with an interpolating curve (such as a polynomial) that is more readily differentiated or integrated. For example, a step function used to model a square wave might be replaced by a trigonometric approximation to facilitate subsequent analysis. In these instances the modeler desires the interpolating curve to approximate closely the essential characteristics of the function it replaces. This type of interpolation is usually called **approximation** and is typically addressed in introductory numerical analysis courses.

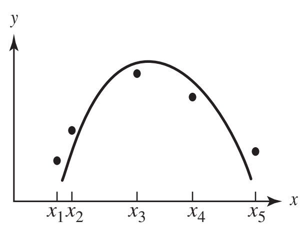


Figure 3.3
Fitting a parabola $y = C_1x^2 + C_2x + C_3$ to the data points

Sources of Error in the Modeling Process

Before discussing criteria on which to base curve-fitting and interpolation decisions, we need to examine the modeling process to ascertain where errors can arise. If error considerations are neglected, undue confidence may be placed in intermediate results, causing faulty decisions in subsequent steps. Our goal is to ensure that all parts of the modeling process are computationally compatible and to consider the effects of cumulative errors likely to exist from previous steps.

For purposes of easy reference, we classify errors under the following category scheme:

1. Formulation error
2. Truncation error
3. Round-off error
4. Measurement error

Formulation errors result from the assumption that certain variables are negligible or from simplifications in describing interrelationships among the variables in the various submodels. For example, when we determined a submodel for braking distance in Chapter 2, we completely neglected road friction, and we assumed a very simple relationship for the nature of the drag force due to air resistance. Formulation errors are present in even the best models.

Truncation errors are attributable to the numerical method used to solve a mathematical problem. For example, we may find it necessary to approximate $\sin x$ with a polynomial representation obtained from the power series

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

An error will be introduced when the series is truncated to produce the polynomial.

Round-off errors are caused by using a finite digit machine for computation. Because all numbers cannot be represented exactly using only finite representations, we must always expect round-off errors to be present. For example, consider a calculator or computer that uses 8-digit arithmetic. Then the number $\frac{1}{3}$ is represented by .33333333 so that 3 times $\frac{1}{3}$ is the number .99999999 rather than the actual value 1. The error 10^{-8} is due to round-off. The ideal real number $\frac{1}{3}$ is an *infinite* string of decimal digits .3333 . . . , but any calculator or computer can do arithmetic only with numbers having finite precision. When many arithmetic operations are performed in succession, each with its own round-off, the accumulated effect of round-off can significantly alter the numbers that are supposed to be the answer. Round-off is just one of the things we have to live with—*and be aware of*—when we use computing machines.

Measurement errors are caused by imprecision in the data collection. This imprecision may include such diverse things as human errors in recording or reporting the data or the actual physical limitations of the laboratory equipment. For example, considerable measurement error would be expected in the data reflecting the reaction distance and the braking distance in the braking distance problem.

3.1 Fitting Models to Data Graphically

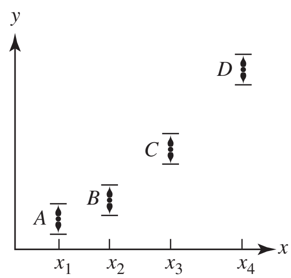
Assume the modeler has made certain assumptions leading to a model of a particular type. The model generally contains one or more parameters, and sufficient data must be gathered to determine them. Let's consider the problem of data collection.

The determination of how many data points to collect involves a trade-off between the cost of obtaining them and the accuracy required of the model. As a minimum, the modeler needs at least as many data points as there are arbitrary constants in the model curve. Additional points are required to determine any arbitrary constants involved with the requirements of the best-fit method we are using. The *range* over which the model is to be used determines the endpoints of the interval for the independent variable(s).

The *spacing* of the data points within that interval is also important because any part of the interval over which the model must fit particularly well can be weighted by using unequal spacing. We may choose to take more data points where maximum use of the model is expected, or we may collect more data points where we anticipate abrupt changes in the dependent variable(s).

Even if the experiment has been carefully designed and the trials meticulously conducted, the modeler needs to appraise the accuracy of the data before attempting to fit the model. How were the data collected? What is the accuracy of the measuring devices used in the collection process? Do any points appear suspicious? Following such an appraisal and elimination (or replacement) of spurious data, it is useful to think of each data point as an interval of relative confidence rather than as a single point. This idea is shown in Figure 3.4. The length of each interval should be commensurate with the appraisal of the errors present in the data collection process.

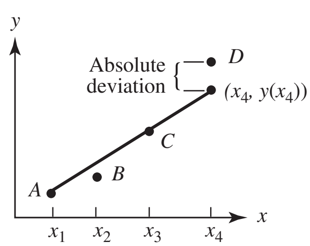
Figure 3.4
Each data point is thought of as an interval of confidence.



Visual Model Fitting with the Original Data

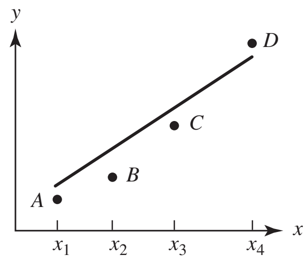
Suppose we want to fit the model $y = ax + b$ to the data shown in Figure 3.4. How might we choose the constants a and b to determine the line that best fits the data? Generally, when more than two data points exist, not all of them can be expected to lie exactly along a single straight line, even if such a line accurately models the relationship between the two variables x and y. Ordinarily, there will be some vertical discrepancy between a few of the data points and any particular line under consideration. We refer to these vertical discrepancies as **absolute deviations** (Figure 3.5). For the best-fitting line, we might try to

Figure 3.5
Minimizing the sum of the absolute deviations from the fitted line



minimize the sum of these absolute deviations, leading to the model depicted in Figure 3.5. Although success may be achieved in minimizing the sum of the absolute deviations, the absolute deviation from individual points may be quite large. For example, consider point D in Figure 3.5. If the modeler has confidence in the accuracy of this data point, there will be concern for the predictions made from the fitted line near the point. As an alternative, suppose a line is selected that minimizes the largest deviation from any point. Applying this criterion to the data points might give the line shown in Figure 3.6.

Figure 3.6
Minimizing the largest absolute deviation from the fitted line



Although these visual methods for fitting a line to data points may appear imprecise, the methods are often quite *compatible* with the accuracy of the modeling process. The grossness of the assumptions and the imprecision involved in the data collection may not warrant a more sophisticated analysis. In such situations, the blind application of one of the analytic methods to be presented in Section 3.2 may lead to models far less appropriate than one obtained graphically. Furthermore, a visual inspection of the model fitted graphically to the data immediately gives an impression of *how good* the fit is and *where* it appears to fit well. Unfortunately, these important considerations are often overlooked in problems with large amounts of data analytically fitted via computer codes. Because the model-fitting portion of the modeling process seems to be more precise and analytic than some of the other steps, there is a tendency to place undue faith in the numerical computations.

Transforming the Data

Most of us are limited visually to fitting only lines. So how can we graphically fit curves as models? Suppose, for example, that a relationship of the form $y = Ce^x$ is suspected for some submodel and the data shown in Table 3.1 have been collected.

The model states that y is proportional to e^x. Thus, if we plot y versus e^x, we should obtain approximately a straight line. The situation is depicted in Figure 3.7. Because the plotted data points do lie approximately along a line that projects through the origin, we conclude that the assumed proportionality is reasonable. From the figure, the slope of the line is approximated as

$$C = \frac{165 - 60.1}{54.6 - 20.1} \approx 3.0$$

Table 3.1 Collected data

x	1	2	3	4
y	8.1	22.1	60.1	165

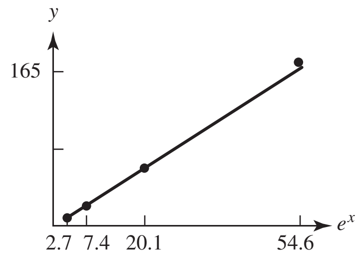


Figure 3.7
Plot of y versus e^x for the data given in Table 3.1

Now let's consider an alternative technique that is useful in a variety of problems. Take the logarithm of each side of the equation $y = Ce^x$ to obtain

$$\ln y = \ln C + x$$

Note that this expression is an equation of a line in the variables $\ln y$ and x. The number $\ln C$ is the intercept when $x = 0$. The transformed data are shown in Table 3.2 and plotted in Figure 3.8. Semilog paper or a computer is useful when plotting large amounts of data.

Table 3.2 The transformed data from Table 3.1

x	1	2	3	4
$\ln y$	2.1	3.1	4.1	5.1

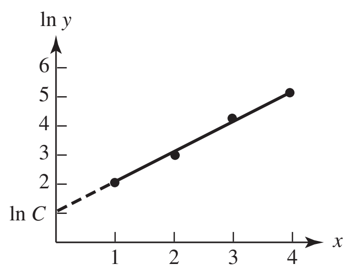


Figure 3.8
Plot of $\ln y$ versus x using Table 3.2

From Figure 3.8, we can determine that the intercept $\ln C$ is approximately 1.1, giving $C = e^{1.1} \approx 3.0$ as before.

A similar transformation can be performed on a variety of other curves to produce linear relationships among the resulting transformed variables. For example, if $y = x^a$, then

$$\ln y = a \ln x$$

is a linear relationship in the transformed variables $\ln y$ and $\ln x$. Here, log–log paper or a computer is useful when plotting large amounts of data.

Let's pause and make an important observation. Suppose we do invoke a transformation and plot $\ln y$ versus x, as in Figure 3.8, and find the line that successfully minimizes the sum of the absolute deviations of the transformed data points. The line then determines $\ln C$, which in turn produces the proportionality constant C. Although it is not obvious, the resulting model $y = Ce^x$ is not the member of the family of exponential curves of the form ke^x that minimizes the sum of the absolute deviations from the original data points (when we plot y versus x). This important idea will be demonstrated both graphically and analytically in the ensuing discussion. When transformations of the form $y = \ln x$ are made, the distance concept is distorted. Although a fit that is compatible with the inherent limitations of a graphical analysis may be obtained, the modeler must be aware of this distortion and *verify the model using the graph from which it is intended to make predictions or conclusions—namely the y versus x graph in the original data rather than the graph of the transformed variables.*

We now present an example illustrating how a transformation may distort distance in the xy-plane. Consider the data plotted in Figure 3.9 and assume the data are expected to fit a model of the form $y = Ce^{1/x}$. Using a logarithmic transformation as before, we find

$$\ln y = \frac{1}{x} + \ln C$$

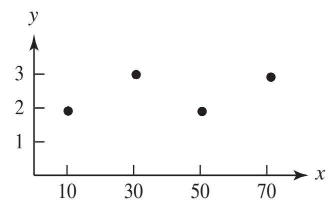


Figure 3.9
A plot of some collected data points

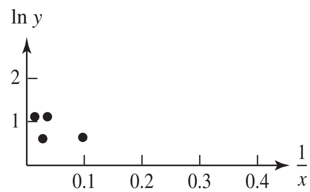


Figure 3.10
A plot of the transformed data points

A plot of the points $\ln y$ versus $1/x$ based on the original data is shown in Figure 3.10. Note from the figure how the transformation distorts the distances between the original data points and squeezes them all together. Consequently, if a straight line is made to fit the transformed data plotted in Figure 3.10, the absolute deviations appear relatively small (i.e., small computed on the Figure 3.10 scale rather than on the Figure 3.9 scale). If we were to plot the fitted model $y = Ce^{1/x}$ to the data in Figure 3.9, we would see that it fits the data relatively poorly, as shown in Figure 3.11.

From the preceding example, it can be seen that if a modeler is not careful when using transformations, he or she can be tricked into selecting a relatively poor model. This realization becomes especially important when comparing alternative models. Very serious errors can be introduced when selecting the best model unless all comparisons are made with the original data (plotted in Figure 3.9 in our example). Otherwise, the choice of best model may be determined by a peculiarity of the transformation rather than on the

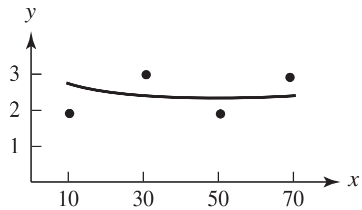


Figure 3.11
A plot of the curve $y = Ce^{1/x}$ based on the value $\ln C = 0.9$ from Figure 3.10

merits of the model and how well it fits the original data. Although the danger of making transformations is evident in this graphical illustration, a modeler may be fooled if he or she is not especially observant because many computer codes fit models by first making a transformation. If the modeler intends to use indicators such as the sum of the absolute deviations to make decisions about the adequacy of a particular submodel or choose among competing submodels, the modeler must first ascertain how those indicators were computed.

3.1 PROBLEMS

1. The model in Figure 3.2 would normally be used to predict behavior between x_1 and x_5. What would be the danger of using the model to predict y for values of x less than x_1 or greater than x_5? Suppose we are modeling the trajectory of a thrown baseball.

2. The following table gives the elongation e in inches per inch (in./in.) for a given stress S on a steel wire measured in pounds per square inch (lb/in.^{2}). Test the model $e = c_1 S$ by plotting the data. Estimate c_1 graphically.

S ($\times 10^{-3}$)	5	10	20	30	40	50	60	70	80	90	100
e ($\times 10^{5}$)	0	19	57	94	134	173	216	256	297	343	390

3. In the following data, x is the diameter of a ponderosa pine in inches measured at breast height and y is a measure of volume—number of board feet divided by 10. Test the model $y = ax^b$ by plotting the transformed data. If the model seems reasonable, estimate the parameters a and b of the model graphically.

x	17	19	20	22	23	25	28	31	32	33	36	37	38	39	41
y	19	25	32	51	57	71	113	141	123	187	192	205	252	259	294

4. In the following data, V represents a mean walking velocity and P represents the population size. We wish to know if we can predict the population size P by observing how fast people walk. Plot the data. What kind of a relationship is suggested? Test the following models by plotting the appropriate transformed data.

   a. $P = aV^b$

   b. $P = a \ln V$

V	2.27	2.76	3.27	3.31	3.70	3.85	4.31	4.39	4.42
P	2500	365	23700	5491	14000	78200	70700	138000	304500

V	4.81	4.90	5.05	5.21	5.62	5.88
P	341948	49375	260200	867023	1340000	1092759

5. The following data represent the growth of a population of fruit flies over a 6-week period. Test the following models by plotting an appropriate set of data. Estimate the parameters of the following models.

a. $P = c_1 t$

b. $P = ae^{bt}$

t (days)	7	14	21	28	35	42
P (number of observed flies)	8	41	133	250	280	297

6. The following data represent (hypothetical) energy consumption normalized to the year 1900. Plot the data. Test the model $Q = ae^{bx}$ by plotting the transformed data. Estimate the parameters of the model graphically.

x	Year	Consumption Q
0	1900	1.00
10	1910	2.01
20	1920	4.06
30	1930	8.17
40	1940	16.44
50	1950	33.12
60	1960	66.69
70	1970	134.29
80	1980	270.43
90	1990	544.57
100	2000	1096.63

7. In 1601 the German astronomer Johannes Kepler became director of the Prague Observatory. Kepler had been helping Tycho Brahe in collecting 13 years of observations on the relative motion of the planet Mars. By 1609 Kepler had formulated his first two laws:

i. Each planet moves on an ellipse with the sun at one focus.

ii. For each planet, the line from the sun to the planet sweeps out equal areas in equal times.

Kepler spent many years verifying these laws and formulating a third law, which relates the planets' orbital periods and mean distances from the sun.

a. Plot the period time T versus the mean distance r using the following updated observational data.

Planet	Period (days)	Mean distance from the sun (millions of kilometers)
Mercury	88	57.9
Venus	225	108.2
Earth	365	149.6
Mars	687	227.9
Jupiter	4,329	778.1
Saturn	10,753	1428.2
Uranus	30,660	2837.9
Neptune	60,150	4488.9
Pluto	90,670	5876.7

b. Assuming a relationship of the form

$$T = Cr^a$$

determine the parameters C and a by plotting $\ln T$ versus $\ln r$. Does the model seem reasonable? Try to formulate Kepler's third law.

3.2 Analytic Methods of Model Fitting

In this section we investigate several criteria for fitting curves to a collection of data points. Each criterion suggests a method for selecting the best curve from a given family so that according to the criterion, the curve most accurately represents the data. We also discuss how the various criteria are related.

Chebyshev Approximation Criterion

In the preceding section we graphically fit lines to a given collection of data points. One of the best-fit criteria used was to minimize the largest distance from the line to any corresponding data point. Let's analyze this geometric construction. Given a collection of m data points (x_i, y_i), $i = 1, 2, \dots, m$, fit the collection to the line $y = mx + b$, determined by the parameters a and b, that minimizes the distance between any data point (x_i, y_i) and its corresponding data point on the line $(x_i, ax_i + b)$. That is, minimize the largest absolute deviation $|y_i - y(x_i)|$ over the entire collection of data points. Now let's generalize this criterion.

Given some function type $y = f(x)$ and a collection of m data points (x_i, y_i), minimize the largest absolute deviation $|y_i - f(x_i)|$ over the entire collection. That is, determine the parameters of the function type $y = f(x)$ that minimizes the number

$$\text{Maximum } |y_i - f(x_i)| \qquad i = 1, 2, \ldots, m \tag{3.1}$$

This important criterion is often called the **Chebyshev approximation criterion**. The difficulty with the Chebyshev criterion is that it is often complicated to apply in practice, at least using only elementary calculus. The optimization problems that result from applying the criterion may require advanced mathematical procedures or numerical algorithms necessitating the use of a computer.

Figure 3.12
The line segment AC is divided into two segments, AB and BC.

For example, suppose you want to measure the line segments AB, BC, and AC represented in Figure 3.12. Assume your measurements yield the estimates $AB = 13$, $BC = 7$, and $AC = 19$. As you should expect in any physical measuring process, discrepancy results. In this situation, the values of AB and BC add up to 20 rather than the estimated $AC = 19$. Let's resolve the discrepancy of 1 unit using the Chebyshev criterion. That is, we will assign values to the three line segments in such a way that the largest absolute deviation between any corresponding pair of assigned and observed values is minimized. Assume the same degree of confidence in each measurement so that each measurement has equal weight. In that case, the discrepancy should be distributed equally across each segment, resulting in the predictions $AB = 12\frac{2}{3}$, $BC = 6\frac{2}{3}$, and $AC = 19\frac{1}{3}$. Thus, each absolute deviation is $\frac{1}{3}$. Convince yourself that reducing any of these deviations causes one of the other deviations to increase. (Remember that $AB + BC$ must equal AC.) Let's formulate the problem symbolically.

Let x_1 represent the true value of the length of the segment AB and x_2 the true value for BC. For ease of our presentation, let r_1, r_2, and r_3 represent the discrepancies between the true and measured values as follows:

$$\begin{aligned} x_1 - 13 &= r_1 \text{ (line segment } AB) \\ x_2 - 7 &= r_2 \text{ (line segment } BC) \\ x_1 + x_2 - 19 &= r_3 \text{ (line segment } AC) \end{aligned}$$

The numbers r_1, r_2, and r_3 are called **residuals**. Note that residuals can be positive or negative, whereas absolute deviations are always positive.

If the Chebyshev approximation criterion is applied, values are assigned to x_1 and x_2 in such a way as to minimize the largest of the three numbers $|r_1|$, $|r_2|$, $|r_3|$. If we call that largest number r, then we want to

$$\text{Minimize } r$$

subject to the three conditions:

$$\begin{aligned} |r_1| \le r \quad &\text{or} \quad -r \le r_1 \le r \\ |r_2| \le r \quad &\text{or} \quad -r \le r_2 \le r \\ |r_3| \le r \quad &\text{or} \quad -r \le r_3 \le r \end{aligned}$$

Each of these conditions can be replaced by two inequalities. For example, $|r_1| \leq r$ can be replaced by $r - r_1 \geq 0$ and $r + r_1 \geq 0$. If this is done for each condition, the problem can be stated in the form of a classical mathematical problem:

$$\text{Minimize } r$$

Subject to

$$\begin{aligned}
r - x_1 \qquad\quad + 13 &\geq 0 \quad (r - r_1 \geq 0)\\
r + x_1 \qquad\quad - 13 &\geq 0 \quad (r + r_1 \geq 0)\\
r \qquad - x_2 + 7 &\geq 0 \quad (r - r_2 \geq 0)\\
r \qquad + x_2 - 7 &\geq 0 \quad (r + r_2 \geq 0)\\
r - x_1 - x_2 + 19 &\geq 0 \quad (r - r_3 \geq 0)\\
r + x_1 + x_2 - 19 &\geq 0 \quad (r + r_3 \geq 0)
\end{aligned}$$

This problem is called a **linear program**. We will discuss linear programs further in Chapter 7. Even large linear programs can be solved by computer implementation of an algorithm known as the **Simplex Method**. In the preceding line segment example, the Simplex Method yields a minimum value of $r = \frac{1}{3}$, and $x_1 = 12\frac{2}{3}$ and $x_2 = 6\frac{2}{3}$.

We now generalize this procedure. Given some function type $y = f(x)$, whose parameters are to be determined, and a collection of m data points (x_i, y_i), define the residuals $r_i = y_i - f(x_i)$. If r represents the largest absolute value of these residuals, then the problem is

$$\text{Minimize } r$$

subject to

$$\left.\begin{aligned} r - r_i &\geq 0\\ r + r_i &\geq 0 \end{aligned}\right\} \quad \text{for } i = 1, 2, \ldots, m$$

Although we discuss linear programs in Chapter 7, we should note here that the model resulting from this procedure is not always a linear program; for example, consider fitting the function $f(x) = \sin kx$. Also note that many computer codes of the Simplex algorithm require using variables that are allowed to assume only nonnegative values. This requirement can be accomplished with simple substitution (see Problem 5).

As we will see, alternative criteria lead to optimization problems that often can be resolved more conveniently. Primarily for this reason, the Chebyshev criterion is not used often for fitting a curve to a finite collection of data points. However, its application should be considered whenever minimizing the largest absolute deviation is important. (We consider several applications of the criterion in Chapter 7.) Furthermore, the principle underlying the Chebyshev criterion is extremely important when one is replacing a function defined over an interval by another function and the largest difference between the two functions over the interval must be minimized. This topic is studied in approximation theory and is typically covered in introductory numerical analysis.

Minimizing the Sum of the Absolute Deviations

When we were graphically fitting lines to the data in Section 3.1, one of our criteria minimized the total sum of the absolute deviations between the data points and their corresponding points on the fitted line. This criterion can be generalized: Given some function type $y = f(x)$ and a collection of m data points (x_i, y_i), minimize the sum of the absolute deviations $|y_i - f(x_i)|$. That is, determine the parameters of the function type $y = f(x)$ to minimize

$$\sum_{i=1}^{m} |y_i - f(x_i)| \tag{3.2}$$

If we let $R_i = |y_i - f(x_i)|$, $i = 1, 2, \ldots, m$ represent each absolute deviation, then the preceding criterion (3.2) can be interpreted as minimizing the length of the line formed by adding together the numbers R_i. This is illustrated for the case $m = 2$ in Figure 3.13.

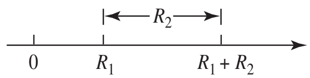


Figure 3.13
A geometric interpretation of minimizing the sum of the absolute deviations

Although we geometrically applied this criterion in Section 3.1 when the function type $y = f(x)$ was a line, the general criterion presents severe problems. To solve this optimization problem using the calculus, we need to differentiate the sum (3.2) with respect to the parameters of $f(x)$ to find the critical points. However, the various derivatives of the sum fail to be continuous because of the presence of the absolute values, so we will not pursue this criterion further. In Chapter 7 we consider other applications of the criterion and present techniques for approximating solutions numerically.

Least-Squares Criterion

Currently, the most frequently used curve-fitting criterion is the **least-squares criterion**. If we use the same notation shown earlier, the problem is to determine the parameters of the function type $y = f(x)$ to minimize the sum

$$\sum_{i=1}^{m} |y_i - f(x_i)|^2 \tag{3.3}$$

Part of the popularity of this criterion stems from the ease with which the resulting optimization problem can be solved using only the calculus of several variables. However, relatively recent advances in mathematical programming techniques (such as the Simplex Method for solving many applications of the Chebyshev criterion) and advances in numerical methods for approximating solutions to the criterion (3.2) promise to dissipate this advantage. The justification for the use of the least-squares method increases when we consider probabilistic arguments that assume the errors are distributed randomly. However, we will not discuss such probabilistic arguments until later in the text.

We now give a geometric interpretation of the least-squares criterion. Consider the case of three data points and let $R_i = |y_i - f(x_i)|$ denote the absolute deviation between the observed and predicted values for $i = 1, 2, 3$. Think of the R_i as the scalar components of a deviation vector, as depicted in Figure 3.14. Thus the vector $\mathbf{R} = R_1\mathbf{i} + R_2\mathbf{j} + R_3\mathbf{k}$ represents the resultant deviation between the observed and predicted values. The magnitude

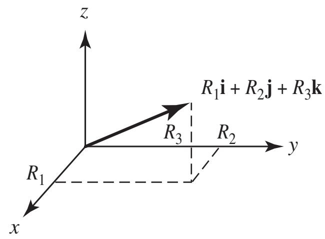


Figure 3.14
A geometric interpretation of the least-squares criterion

of the deviation vector is given by

$$|\mathbf{R}| = \sqrt{R_1^2 + R_2^2 + R_3^2}$$

To minimize $|\mathbf{R}|$ we can minimize $|\mathbf{R}|^2$ (see Problem 1). Thus, the least-squares problem is to determine the parameters of the function type $y = f(x)$ such that

$$|\mathbf{R}|^2 = \sum_{i=1}^{3} R_i^2 = \sum_{i=1}^{3} |y_i - f(x_i)|^2$$

is minimized. That is, we may interpret the least-squares criterion as minimizing the magnitude of the vector whose coordinates represent the absolute deviation between the observed and predicted values.

Relating the Criteria

The geometric interpretations of the three curve-fitting criteria help in providing a qualitative description comparing the criteria. Minimizing the sum of the absolute deviations tends to treat each data point with equal weight and to average the deviations. The Chebyshev criterion gives more weight to a single point potentially having a large deviation. The least-squares criterion is somewhere in between as far as weighting individual points with significant deviations is concerned. But let's be more precise. Because the Chebyshev and least-squares criteria are the most convenient to apply analytically, we now derive a method for relating the deviations resulting from using these two criteria.

Suppose the Chebyshev criterion is applied and the resulting optimization problem solved to yield the function $f_1(x)$. The absolute deviations resulting from the fit are defined as follows:

$$|y_i - f_1(x_i)| = c_i, \quad i = 1, 2, \ldots, m$$

Now define $c_{\max}$ as the largest of the absolute deviations c_i. There is a special significance attached to $c_{\max}$. Because the parameters of the function $f_1(x)$ are determined so as to minimize the value of $c_{\max}$, it is the minimal largest absolute deviation obtainable.

On the other hand, suppose the least-squares criterion is applied and the resulting optimization problem solved to yield the function $f_2(x)$. The absolute deviations resulting from the fit are then given by

$$|y_i - f_2(x_i)| = d_i, \quad i = 1, 2, \ldots, m$$

Define $d_{\max}$ as the largest of the absolute deviations d_i. At this point it can only be said that $d_{\max}$ is at least as large as $c_{\max}$ because of the special significance of the latter as previously discussed. However, let's attempt to relate $d_{\max}$ and $c_{\max}$ more precisely.

The special significance the least-squares criterion attaches to the d_i is that the sum of their squares is the smallest such sum obtainable. Thus, it must be true that

$$d_1^2 + d_2^2 + \cdots + d_m^2 \leq c_1^2 + c_2^2 + \cdots + c_m^2$$

Because $c_i \leq c_{\max}$ for every i, these inequalities imply

$$d_1^2 + d_2^2 + \cdots + d_m^2 \leq mc_{\max}^2$$

or

$$\sqrt{\frac{d_1^2 + d_2^2 + \cdots + d_m^2}{m}} \leq c_{\max}$$

For ease of discussion, define

$$D = \sqrt{\frac{d_1^2 + d_2^2 + \cdots + d_m^2}{m}}$$

Thus,

$$D \leq c_{\max} \leq d_{\max}$$

This last relationship is very revealing. Suppose it is more convenient to apply the least-squares criterion in a particular situation, but there is concern about the largest absolute deviation $c_{\max}$ that may result. If we compute D, a lower bound on $c_{\max}$ is obtained, and $d_{\max}$ gives an upper bound. Thus, if there is considerable difference between D and $d_{\max}$, the modeler should consider applying the Chebyshev criterion.

3.2 PROBLEMS

1. Using elementary calculus, show that the minimum and maximum points for $y = f(x)$ occur among the minimum and maximum points for $y = f^2(x)$. Assuming $f(x) \geq 0$, why can we minimize $f(x)$ by minimizing $f^2(x)$?

2. For each of the following data sets, formulate the mathematical model that minimizes the largest deviation between the data and the line $y = ax + b$. If a computer is available, solve for the estimates of a and b.

a.

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

b.

x	29.1	48.2	72.7	92.0	118	140	165	199
y	0.0493	0.0821	0.123	0.154	0.197	0.234	0.274	0.328

c.

x	2.5	3.0	3.5	4.0	4.5	5.0	5.5
y	4.32	4.83	5.27	5.74	6.26	6.79	7.23

3. For the following data, formulate the mathematical model that minimizes the largest deviation between the data and the model $y = c_1x^2 + c_2x + c_3$. If a computer code is available, solve for the estimates of c_1, c_2, and c_3.

x	0.1	0.2	0.3	0.4	0.5
y	0.06	0.12	0.36	0.65	0.95

4. For the following data, formulate the mathematical model that minimizes the largest deviation between the data and the model $P = ae^{bt}$. If a computer code is available, solve for the estimates of a and b.

t	7	14	21	28	35	42
P	8	41	133	250	280	297

5. Suppose the variable x_1 can assume any real value. Show that the following substitution using nonnegative variables x_2 and x_3 permits x_1 to assume any real value.

$$x_1 = x_2 - x_3, \qquad \text{where } x_1 \text{ is unconstrained}$$

and

$$x_2 \geq 0 \qquad \text{and} \qquad x_3 \geq 0$$

Thus, if a computer code allows only nonnegative variables, the substitution allows for solving the linear program in the variables x_2 and x_3 and then recovering the value of the variable x_1.

3.3 Applying the Least-Squares Criterion

Suppose that our assumptions lead us to expect a model of a certain type and that data have been collected and analyzed. In this section the least-squares criterion is applied to estimate the parameters for several types of curves.

Fitting a Straight Line

Suppose a model of the form $y = Ax + B$ is expected and it has been decided to use the m data points (x_i, y_i), $i = 1, 2, \ldots, m$ to estimate A and B. Denote the least-squares estimate of $y = Ax + B$ by $y = ax + b$. Applying the least-squares criterion (3.3) to this situation requires the minimization of

$$S = \sum_{i=1}^{m} [y_i - f(x_i)]^2 = \sum_{i=1}^{m} (y_i - ax_i - b)^2$$

A necessary condition for optimality is that the two partial derivatives $\partial S/\partial a$ and $\partial S/\partial b$ equal zero, yielding the equations

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{m}(y_i - ax_i - b)x_i = 0$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{m}(y_i - ax_i - b) = 0$$

These equations can be rewritten to give

$$\left.\begin{aligned} a\sum_{i=1}^{m} x_i^2 + b\sum_{i=1}^{m} x_i &= \sum_{i=1}^{m} x_i y_i \\ a\sum_{i=1}^{m} x_i + mb &= \sum_{i=1}^{m} y_i \end{aligned}\right\} \tag{3.4}$$

The preceding equations can be solved for a and b once all the values for x_i and y_i are substituted into them. The solutions (see Problem 1 at the end of this section) for the parameters a and b are easily obtained by elimination and are found to be

$$a = \frac{m\sum x_i y_i - \sum x_i \sum y_i}{m\sum x_i^2 - \left(\sum x_i\right)^2}, \qquad \text{the } \textbf{slope} \tag{3.5}$$

and

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i}{m\sum x_i^2 - \left(\sum x_i\right)^2}, \qquad \text{the } \textbf{intercept} \tag{3.6}$$

Computer codes are easily written to compute these values for a and b for any collection of data points. Equations (3.4) are called the **normal equations**.

Fitting a Power Curve

Now let's use the least-squares criterion to fit a curve of the form $y = Ax^n$, where n is fixed, to a given collection of data points. Call the least-squares estimate of the model $f(x) = ax^n$. Application of the criterion then requires minimization of

$$S = \sum_{i=1}^{m} [y_i - f(x_i)]^2 = \sum_{i=1}^{m} \left[y_i - ax_i^n\right]^2$$

A necessary condition for optimality is that the derivative ds/da equal zero, giving the equation

$$\frac{dS}{da} = -2\sum_{i=1}^{m} x_i^n \left[y_i - ax_i^n\right] = 0$$

Solving the equation for a yields

$$a = \frac{\sum x_i^n y_i}{\sum x_i^{2n}} \tag{3.7}$$

Remember, the number n is *fixed* in Equation (3.7).

The least-squares criterion can be applied to other models as well. The limitation in applying the method lies in calculating the various derivatives required in the optimization process, setting these derivatives to zero, and solving the resulting equations for the parameters in the model type.

For example, let's fit $y = Ax^2$ to the data shown in Table 3.3 and predict the value of y when $x = 2.25$.

Table 3.3 Data collected to fit $y = Ax^2$

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

In this case, the least-squares estimate a is given by

$$a = \frac{\sum x_i^2 y_i}{\sum x_i^4}$$

We compute $\sum x_i^4 = 61.1875$, $\sum x_i^2 y_i = 195.0$ to yield $a = 3.1869$ (to four decimal places). This computation gives the least-squares approximate model

$$y = 3.1869x^2$$

When $x = 2.25$, the predicted value for y is 16.1337.

Transformed Least-Squares Fit

Although the least-squares criterion appears easy to apply in theory, in practice it may be difficult. For example, consider fitting the model $y = Ae^{Bx}$ using the least-squares criterion. Call the least-squares estimate of the model $f(x) = ae^{bx}$. Application of the criterion then requires the minimization of

$$S = \sum_{i=1}^{m} [y_i - f(x_i)]^2 = \sum_{i=1}^{m} \left[y_i - ae^{bx_i}\right]^2$$

A necessary condition for optimality is that $\partial S/\partial a = \partial S/\partial b = 0$. Formulate the conditions and convince yourself that solving the resulting system of nonlinear equations would not be easy. Many simple models result in derivatives that are very complex or in systems of equations that are difficult to solve. For this reason, we use transformations that allow us to *approximate* the least-squares model.

In graphically fitting lines to data in Section 3.1, we often found it convenient to transform the data first and then fit a line to the transformed points. For example, in graphically fitting $y = Ce^x$, we found it convenient to plot $\ln y$ versus x and then fit a line to the transformed data. The same idea can be used with the least-squares criterion to simplify the computational aspects of the process. In particular, if a convenient substitution can be found so that the problem takes the form $Y = AX + B$ in the transformed variables X and Y, then Equation (3.4) can be used to fit a line to the transformed variables. We illustrate the technique with the example that we just worked out.

Suppose we wish to fit the power curve $y = Ax^N$ to a collection of data points. Let's denote the estimate of A by α and the estimate of N by n. Taking the logarithm of both sides of the equation $y = \alpha x^n$ yields

$$\ln y = \ln \alpha + n \ln x \tag{3.8}$$

When the variables $\ln y$ versus $\ln x$ are plotted, Equation (3.8) yields a straight line. On that graph, $\ln \alpha$ is the intercept when $\ln x = 0$ and the slope of the line is n. Using Equations (3.5) and (3.6) to solve for the slope n and intercept $\ln \alpha$ with the transformed variables and $m = 5$ data points, we have

$$n = \frac{5 \sum (\ln x_i)(\ln y_i) - \left(\sum \ln x_i\right)\left(\sum \ln y_i\right)}{5 \sum (\ln x_i)^2 - \left(\sum \ln x_i\right)^2}$$

$$\ln \alpha = \frac{\sum (\ln x_i)^2 (\ln y_i) - \left(\sum \ln x_i\right)(\ln y_i) \sum \ln x_i}{5 \sum (\ln x_i)^2 - \left(\sum \ln x_i\right)^2}$$

For the data displayed in Table 3.3 we get $\sum \ln x_i = 1.3217558$, $\sum \ln y_i = 8.359597801$, $\sum (\ln x_i)^2 = 1.9648967$, $\sum (\ln x_i)(\ln y_i) = 5.542315175$, yielding $n = 2.062809314$ and $\ln \alpha = 1.126613508$, or $\alpha = 3.085190815$. Thus, our least-squares best fit of Equation (3.8) is (rounded to four decimal places)

$$y = 3.0852x^{2.0628}$$

This model predicts $y = 16.4348$ when $x = 2.25$. Note, however, that this model fails to be a quadratic like the one we fit previously.

Suppose we still wish to fit a *quadratic* $y = Ax^2$ to the collection of data. Denote the estimate of A by a_1 to distinguish this constant from the constants a and α computed previously. Taking the logarithm of both sides of the equation $y = a_1 x^2$ yields

$$\ln y = \ln a_1 + 2 \ln x$$

In this situation the graph of $\ln y$ versus $\ln x$ is a straight line of slope 2 and intercept $\ln a_1$. Using the second equation in (3.4) to compute the intercept, we have

$$2 \sum \ln x_i + 5 \ln a_1 = \sum \ln y_i$$

For the data displayed in Table 3.3, we get $\sum \ln x_i = 1.3217558$ and $\sum \ln y_i = 8.359597801$. Therefore, this last equation gives $\ln a_1 = 1.14321724$, or $a_1 = 3.136844129$, yielding the least-squares best fit (rounded to four decimal places)

$$y = 3.1368x^2$$

The model predicts $y = 15.8801$ when $x = 2.25$, which differs significantly from the value 16.1337 predicted by the first quadratic $y = 3.1869x^2$ obtained as the least-squares best fit of $y = Ax^2$ without transforming the data. We compare these two quadratic models (as well as a third model) in the next section.

The preceding example illustrates two facts. First, if an equation can be transformed to yield an equation of a straight line in the transformed variables, Equation (3.4) can be used directly to solve for the slope and intercept of the transformed graph. Second, the least-squares best fit to the transformed equations *does not* coincide with the least-squares best fit of the original equations. The reason for this discrepancy is that the resulting optimization problems are different. In the case of the original problem, we are finding the curve that minimizes the sum of the squares of the deviations using the original data, whereas in the case of the transformed problem, we are minimizing the sum of the squares of the deviations using the *transformed* variables.

3.3 PROBLEMS

1. Solve the two equations given by (3.4) to obtain the values of the parameters given by Equations (3.5) and (3.6), respectively.

2. Use Equations (3.5) and (3.6) to estimate the coefficients of the line $y = ax + b$ such that the sum of the squared deviations between the line and the following data points is minimized.

a.

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

b.

x	29.1	48.2	72.7	92.0	118	140	165	199
y	0.0493	0.0821	0.123	0.154	0.197	0.234	0.274	0.328

c.

x	2.5	3.0	3.5	4.0	4.5	5.0	5.5
y	4.32	4.83	5.27	5.74	6.26	6.79	7.23

For each problem, compute D and $d_{\max}$ to bound $c_{\max}$. Compare the results to your solutions to Problem 2 in Section 3.2.

3. Derive the equations that minimize the sum of the squared deviations between a set of data points and the quadratic model $y = c_1x^2 + c_2x + c_3$. Use the equations to find estimates of c_1, c_2, and c_3 for the following set of data.

x	0.1	0.2	0.3	0.4	0.5
y	0.06	0.12	0.36	0.65	0.95

Compute D and $d_{\max}$ to bound $c_{\max}$. Compare the results with your solution to Problem 3 in Section 3.2.

4. Make an appropriate transformation to fit the model $P = ae^{bt}$ using Equation (3.4). Estimate a and b.

t	7	14	21	28	35	42
P	8	41	133	250	280	297

5. Examine closely the system of equations that result when you fit the quadratic in Problem 3. Suppose $c_2 = 0$. What would be the corresponding system of equations? Repeat for the cases $c_1 = 0$ and $c_3 = 0$. Suggest a system of equations for a cubic. Check your result. Explain how you would generalize the system of Equation (3.4) to fit any polynomial. Explain what you would do if one or more of the coefficients in the polynomial were zero.

6. A general rule for computing a person's weight is as follows: For a female, multiply the height in inches by 3.5 and subtract 108; for a male, multiply the height in inches by 4.0 and subtract 128. If the person is small bone-structured, adjust this computation by subtracting 10%; for a large bone-structured person, add 10%. No adjustment is made for an average-size person. Gather data on the weight versus height of people of differing age, size, and gender. Using Equation (3.4), fit a straight line to your data for males and another straight line to your data for females. What are the slopes and intercepts of those lines? How do the results compare with the general rule?

In Problems 7–10, fit the data with the models given, using least squares.

7.

x	1	2	3	4	5
y	1	1	2	2	4

a. $y = b + ax$
b. $y = ax^2$

8. Data for stretch of a spring

$x(\times 10^{-3})$	5	10	20	30	40	50	60	70	80	90	100
$y(\times 10^{-5})$	0	19	57	94	134	173	216	256	297	343	390

a. $y = ax$
b. $y = b + ax$
c. $y = ax^2$

9. Data for the ponderosa pine

x	17	19	20	22	23	25	28	31	32	33	36	37	39	42
y	19	25	32	51	57	71	113	140	153	187	192	205	250	260

a. $y = ax + b$
b. $y = ax^2$

c. $y = ax^3$

d. $y = ax^3 + bx^2 + c$

10. Data for planets

Body	Period (sec)	Distance from sun (m)
Mercury	7.60×10^6	5.79×10^{10}
Venus	1.94×10^7	1.08×10^{11}
Earth	3.16×10^7	1.5×10^{11}
Mars	5.94×10^7	2.28×10^{11}
Jupiter	3.74×10^8	7.79×10^{11}
Saturn	9.35×10^8	1.43×10^{12}
Uranus	2.64×10^9	2.87×10^{12}
Neptune	5.22×10^9	4.5×10^{12}
Pluto	7.82×10^9	5.91×10^{12}

Fit the model $y = ax^{3/2}$.

3.3 PROJECTS

1. Compete the requirements of the module “Curve Fitting via the Criterion of Least Squares,” by John W. Alexander, Jr., UMAP 321. (See enclosed CD for UMAP module.) This unit provides an easy introduction to correlations, scatter diagrams (polynomial, logarithmic, and exponential scatters), and lines and curves of regression. Students construct scatter diagrams, choose appropriate functions to fit specific data, and use a computer program to fit curves. Recommended for students who wish an introduction to statistical measures of correlation.

2. Select a project from Projects 1–7 in Section 2.3 and use least squares to fit your proposed proportionality model. Compare your least-squares results with the model used from Section 2.3. Find the bounds on the Chebyshev criterion and interpret the results.

3.3 Further Reading

Burden, Richard L., & J. Douglas Faires. *Numerical Analysis*, 7th ed. Pacific Grove, CA: Brooks/Cole, 2001.

Cheney, E. Ward, & David Kincaid. *Numerical Mathematics and Computing*. Monterey, CA: Brooks/Cole, 1984.

Cheney, E. Ward, & David Kincaid. *Numerical Analysis*, 4th ed. Pacific Grove, CA: Brooks/Cole, 1999.

Hamming, R. W. *Numerical Methods for Scientists and Engineers*. New York: McGraw-Hill, 1973.

Stiefel, Edward L. *An Introduction to Numerical Mathematics*. New York: Academic Press, 1963.

3.4 Choosing a Best Model

Let's consider the adequacy of the various models of the form $y = Ax^2$ that we fit using the least-squares and transformed least-squares criteria in the previous section. Using the least-squares criterion, we obtained the model $y = 3.1869x^2$. One way of evaluating how well the model fits the data is to compute the deviations between the model and the actual data. If we compute the sum of the squares of the deviations, we can bound $c_{\max}$ as well. For the model $y = 3.1869x^2$ and the data given in Table 3.3, we compute the deviations shown in Table 3.4.

Table 3.4 Deviations between the data in Table 3.3 and the fitted model $y = 3.1869x^2$

x_i	0.5	1.0	1.5	2.0	2.5
y_i	0.7	3.4	7.2	12.4	20.1
$y_i - y(x_i)$	−0.0967	0.2131	0.02998	−0.3476	0.181875

From Table 3.4 we compute the sum of the squares of the deviations as 0.20954, so $D = (0.20954/5)^{1/2} = 0.204714$. Because the largest absolute deviation is 0.3476 when $x = 2.0$, $c_{\max}$ can be bounded as follows:

$$D = 0.204714 \leq c_{\max} \leq 0.3476 = d_{\max}$$

Let's find $c_{\max}$. Because there are five data points, the mathematical problem is to minimize the largest of the five numbers $|r_i| = |y_i - y(x_i)|$. Calling that largest number r, we want to minimize r subject to $r \geq r_i$ and $r \geq -r_i$ for each $i = 1, 2, 3, 4, 5$. Denote our model by $y(x) = a_2x^2$. Then, substitution of the observed data points in Table 3.3 into the inequalities $r \geq r_i$ and $r \geq -r_i$ for each $i = 1, 2, 3, 4, 5$ yields the following linear program:

$$\text{Minimize } r$$

subject to

$$\begin{aligned}
r - r_1 &= r - (0.7 - 0.25a_2) \geq 0 \\
r + r_1 &= r + (0.7 - 0.25a_2) \geq 0 \\
r - r_2 &= r - (3.4 - a_2) \geq 0 \\
r + r_2 &= r + (3.4 - a_2) \geq 0 \\
r - r_3 &= r - (7.2 - 2.25a_2) \geq 0 \\
r + r_3 &= r + (7.2 - 2.25a_2) \geq 0 \\
r - r_4 &= r - (12.4 - 4a_2) \geq 0 \\
r + r_4 &= r + (12.4 - 4a_2) \geq 0 \\
r - r_5 &= r - (20.1 - 6.25a_2) \geq 0 \\
r + r_5 &= r + (20.1 - 6.25a_2) \geq 0
\end{aligned}$$

In Chapter 7 we show that the solution of the preceding linear program yields $r = 0.28293$ and $a_2 = 3.17073$. Thus, we have reduced our largest deviation from $d_{\max} = 0.3476$ to $c_{\max} = 0.28293$. Note that we can reduce the largest deviation no further than 0.28293 for the model type $y = Ax^2$.

We have now determined three estimates of the parameter A for the model type $y = Ax^2$. Which estimate is best? For each model we can readily compute the deviations from each data point as recorded in Table 3.5.

Table 3.5 Summary of the deviations for each model $y = Ax^2$

x_i	y_i	$y_i - 3.1869x_i^2$	$y_i - 3.1368x_i^2$	$y_i - 3.17073x_i^2$
0.5	0.7	−0.0967	−0.0842	−0.0927
1.0	3.4	0.2131	0.2632	0.2293
1.5	7.2	0.029475	0.1422	0.0659
2.0	12.4	−0.3476	−0.1472	−0.2829
2.5	20.1	0.181875	0.4950	0.28293

For each of the three models we can compute the sum of the squares of the deviations and the maximum absolute deviation. The results are shown in Table 3.6.

As we would expect, each model has something to commend it. However, notice the increase in the sum of the squares of the deviations in the transformed least-squares model. It is tempting to apply a simple rule, such as choose the model with the smallest absolute deviation. (Other statistical indicators of goodness of fit exist as well. For example, see *Probability and Statistics in Engineering and Management Science*, by William W. Hines and Douglas C. Montgomery, New York: Wiley, 1972.) These indicators are useful for eliminating obviously poor models, but there is no easy answer to the question, Which model is best? The model with the smallest absolute deviation or the smallest sum of squares may fit very poorly over the range where you intend to use it most. Furthermore, as you will see in Chapter 4, models can easily be constructed that pass through each data point, thereby yielding a zero sum of squares and zero maximum deviation. So we need to answer the question of which model is best on a case-by-case basis, taking into account such things as the purpose of the model, the precision demanded by the scenario, the accuracy of the data, and the range of values for the independent variable over which the model will be used.

When choosing among models or judging the adequacy of a model, we may find it tempting to rely on the value of the best-fit criterion being used. For example, it is tempting to choose the model that has the smallest sum of squared deviations for the given data set or to conclude that a sum of squared deviations less than a predetermined value indicates

Table 3.6 Summary of the results for the three models

Criterion	Model	$\sum [y_i - y(x_i)]^2$	Max $\lvert y_i - y(x_i)\rvert$
Least-squares	$y = 3.1869x^2$	0.2095	0.3476
Transformed least-squares	$y = 3.1368x^2$	0.3633	0.4950
Chebyshev	$y = 3.17073x^2$	0.2256	0.28293

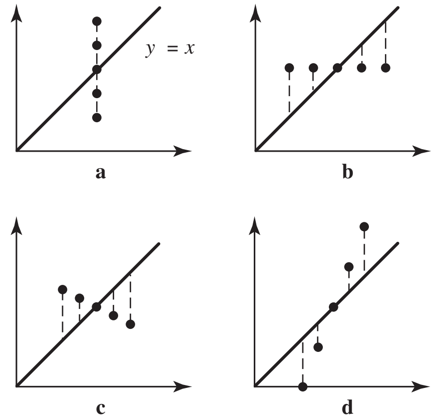


Figure 3.15
In all of these graphs, the model $y = x$ has the same sum of squared deviations.

a good fit. However, in isolation these indicators may be very misleading. For example, consider the data displayed in Figure 3.15. In all of the four cases, the model $y = x$ results in exactly the same sum of squared deviations. Without the benefit of the graphs, therefore, we might conclude that in each case the model fits the data about the same. However, as the graphs show, there is a significant variation in each model's ability to capture the trend of the data. The following examples illustrate how the various indicators may be used to help in reaching a decision on the adequacy of a particular model. Normally, a graphical plot is of great benefit.

EXAMPLE 1 *Vehicular Stopping Distance*

Let's reconsider the problem of predicting a motor vehicle's stopping distance as a function of its speed. (This problem was addressed in Sections 2.2 and 3.3.) In Section 3.3 the submodel in which reaction distance d_r was proportional to the velocity v was tested graphically, and the constant of proportionality was estimated to be 1.1. Similarly, the submodel predicting a proportionality between braking distance d_b and the square of the velocity was tested. We found reasonable agreement with the submodel and estimated the proportionality constant to be 0.054. Hence, the model for stopping distance was given by

$$d = 1.1v + 0.054v^2 \tag{3.9}$$

We now fit these submodels analytically and compare the various fits.

To fit the model using the least-squares criterion, we use the formula from Equation (3.7):

$$A = \frac{\sum x_i y_i}{\sum x_i^2}$$

where y_i denotes the driver reaction distance and x_i denotes the speed at each data point. For the 13 data points given in Table 2.3, we compute $\sum x_i y_i = 40905$ and $\sum x_i^2 = 37050$, giving $A = 1.104049$.

For the model type $d_b = Bv^2$, we use the formula

$$B = \frac{\sum x_i^2 y_i}{\sum x_i^4}$$

where y_i denotes the average braking distance and x_i denotes the speed at each data point. For the 13 data points given in Table 2.4, we compute $\sum x_i^2 y_i = 8258350$ and $\sum x_i^4 = 152343750$, giving $B = 0.054209$. Because the data are relatively imprecise and the modeling is done qualitatively, we round the coefficients to obtain the model

$$d = 1.104v + 0.0542v^2 \tag{3.10}$$

Model (3.10) does not differ significantly from that obtained graphically in Chapter 3. Next, let's analyze how well the model fits. We can readily compute the deviations between the observed data points in Table 2.3 and the values predicted by Models (3.9) and (3.10). The deviations are summarized in Table 3.7. The fits of both models are very similar. The largest absolute deviation for Model (3.9) is 30.4 and for Model (3.10) it is 28.8. Note that both models overestimate the stopping distance up to 70 mph, and then they begin to underestimate the stopping distance. A better-fitting model would be obtained by directly fitting the data for total stopping distance to

$$d = k_1 v + k_2 v^2$$

instead of fitting the submodels individually as we did. The advantage of fitting the submodels individually and then testing each submodel is that we can measure how well they explain the behavior.

A plot of the proposed model(s) and the observed data points is useful to determine how well the model fits the data. Model (3.10) and the observations are plotted in Figure 3.16. It

Table 3.7 Deviations from the observed data points and Models (3.9) and (3.10)

Speed	Graphical model (3.9)	Least-squares model (3.10)
20	1.6	1.76
25	5.25	5.475
30	8.1	8.4
35	13.15	13.535
40	14.4	14.88
45	16.35	16.935
50	17	17.7
55	14.35	15.175
60	12.4	13.36
65	7.15	8.255
70	−1.4	−0.14
75	−14.75	−13.325
80	−30.4	−28.8

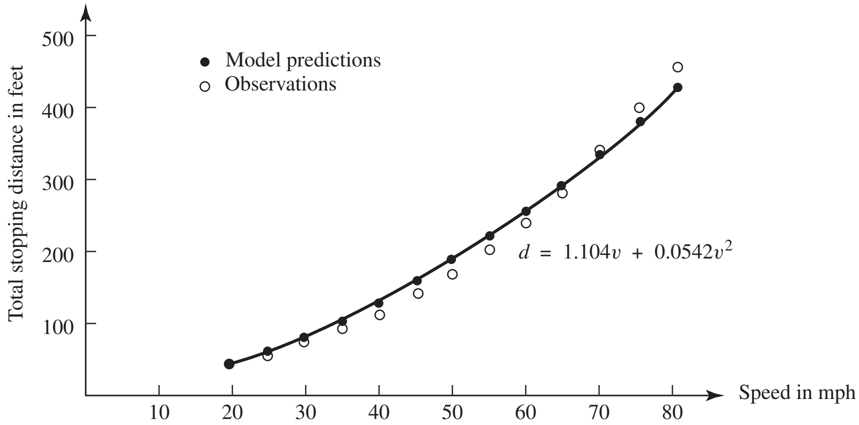


Figure 3.16

A plot of the proposed model and the observed data points provides a visual check on the adequacy of the model.

is evident from the figure that a definite trend exists in the data and that Model (3.10) does a reasonable job of capturing that trend, especially at the lower speeds.

A powerful technique for quickly determining where the model is breaking down is to plot the deviations (residuals) as a function of the independent variable(s). For Model (3.10), a plot of the deviations is given in Figure 3.17 showing that the model is indeed reasonable up to 70 mph. Beyond 70 mph there is a breakdown in the model's ability to predict the observed behavior.

Let's examine Figure 3.17 more closely. Note that although the deviations up to 70 mph are relatively small, they are all positive. If the model fully explains the behavior, not only

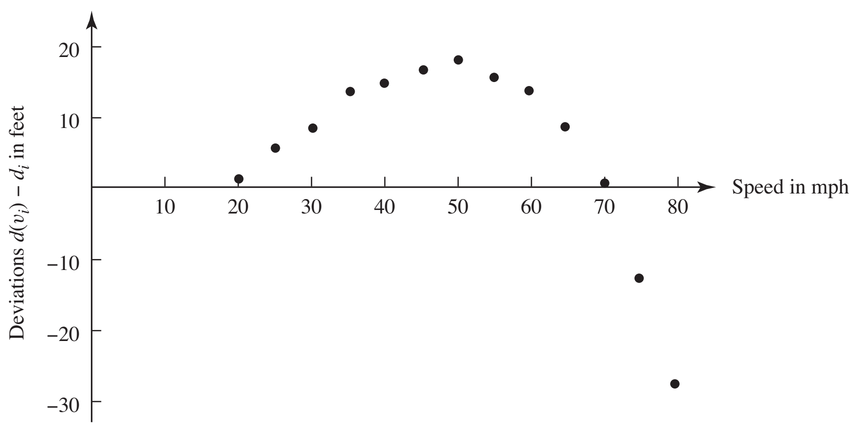


Figure 3.17

A plot of the deviations (residuals) reveals those regions where the model does not fit well.

should the deviations be small, but some should be positive and some negative. Why? In Figure 3.17 we note a definite pattern in the nature of the deviations, which might cause us to reexamine the model and/or the data. The nature of the pattern in the deviations can give us clues on how to refine the model further. In this case, the imprecision in the data collection process probably does not warrant further model refinement.

3.4 PROBLEMS

For Problems 1–6, find a model using the least-squares criterion either on the data or on the transformed data (as appropriate). Compare your results with the graphical fits obtained in the problem set 3.1 by computing the deviations, the maximum absolute deviation, and the sum of the squared deviations for each model. Find a bound on c_{max} if the model was fit using the least-squares criterion.

1. Problem 3 in Section 3.1

2. Problem 4a in Section 3.1

3. Problem 4b in Section 3.1

4. Problem 5a in Section 3.1

5. Problem 2 in Section 3.1

6. Problem 6 in Section 3.1

7. a. In the following data, W represents the weight of a fish (bass) and l represents its length. Fit the model $W = kl^3$ to the data using the least-squares criterion.

Length, l (in.)	14.5	12.5	17.25	14.5	12.625	17.75	14.125	12.625
Weight, W (oz)	27	17	41	26	17	49	23	16

b. In the following data, g represents the girth of a fish. Fit the model $W = klg^2$ to the data using the least-squares criterion.

Length, l (in.)	14.5	12.5	17.25	14.5	12.625	17.75	14.125	12.625
Girth, g (in.)	9.75	8.375	11.0	9.75	8.5	12.5	9.0	8.5
Weight, W (oz)	27	17	41	26	17	49	23	16

c. Which of the two models fits the data better? Justify fully. Which model do you prefer? Why?

8. Use the data presented in Problem 7b to fit the models $W = cg^3$ and $W = kgl^2$. Interpret these models. Compute appropriate indicators and determine which model is best. Explain.

3.4 PROJECTS

1. Write a computer program that finds the least-squares estimates of the coefficients in the following models.
   a. $y = ax^2 + bx + c$
   b. $y = ax^n$
2. Write a computer program that computes the deviation from the data points and any model that the user enters. Assuming that the model was fitted using the least-squares criterion, compute D and $d_{\max}$. Output each data point, the deviation from each data point, D, $d_{\max}$, and the sum of the squared deviations.
3. Write a computer program that uses Equations (3.4) and the appropriate transformed data to estimate the parameters of the following models.
   a. $y = bx^n$
   b. $y = be^{ax}$
   c. $y = a \ln x + b$
   d. $y = ax^2$
   e. $y = ax^3$