MEMORANDUM
RM-5766-SA
DECEMBER 1968

SOME CURVE-FITTING FUNDAMENTALS

R. L. Petruschell

This research is supported by the Department of Defense, under Contract DAHC15 67 C 0150, monitored by the Assistant Secretary of Defense (Systems Analysis). Views and conclusions expressed herein should not be interpreted as representing the official opinion or policy of SA.

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

This study is presented as a competent treatment of the subject, worthy of publication. The Rand Corporation vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the authors.

Published by The RAND Corporation

PREFACE

This Memorandum was written for the Office of the Assistant Secretary of Defense (OASD), Systems Analysis. It supplements material presented in the proposed OASD handbook, *An Introduction to Military Hardware Cost Analysis*.

The cost analyst draws heavily on formal statistics in general and regression analysis in particular when developing estimating relationships. There are numerous times when, because of data limitations, he must instead rely on mechanical curve fitting and the development of empirical equations.

The material presented here is intended to provide the practicing cost analyst with a basic knowledge of the mechanics of curve fitting and of the properties of the equations he uses. For the most part, the choice of material reflects an attempt to answer those questions which years of experience have shown to be most common and troublesome.

Similar information can be found in other sources, but to the best of the author's knowledge does not exist in any single source. The integration of analytic geometry with curve-fitting methods and the selection of the material itself are the unique features of this presentation. The mathematical discussions are purposely intuitive. They are intended to be understandable by persons having, at best, limited mathematical training.

Special-purpose functional forms and curve-fitting methods have not been included. Only those forms and methods that have already proven to be of general use to the cost analyst are described here.

Although the computational schemes presented in the Memorandum are for the most part suitable for the desk calculator, high-speed digital computers are widely available today, and should be used when possible.

SUMMARY

Much of the difficulty cost analysts have with curve fitting results from an inadequate grounding in the analytic geometry of the empirical equations with which they work. This Memorandum attempts to provide a concise but relatively thorough discussion of this subject while at the same time demonstrating selected methods for doing mechanical curve fitting.

The material is presented in three parts. Section I discusses the properties of the straight line, the exponential, the power function, and the parabola. Included in the discussion of the exponential are the laws of exponents and hence logarithms. Emphasis is on providing insights into the impact of the parameter values on the form of the resultant curves. Graphical illustrations are used extensively.

Section II presents different methods of using these curves to describe the relationship between two variables. It discusses the method of selected points, the method of averages, and the method of least squares, making considerable use of scatter diagrams. It describes a number of measures of goodness of fit including the standard deviation, the coefficient of variation, and an average percent deviation. Throughout this section, computational procedures are carried out in complete detail.

The discussion of curve fitting is continued in Section III, where cases with more than two variables are considered. By using the method of successive approximations, the initial discussion attempts to convey the idea of a net relationship between two variables, eliminating influence of any others, and thus to clarify the meaning of the coefficients in the multivariate linear equation. The method of least squares

is shown to produce the same results as did the method of successive approximations and with significantly less computational effort. The discussion turns next to the nonlinear case. Each of the functional forms described earlier--the exponential, the power function, and the parabola--is used to describe a nonlinear relationship between three variables. Although the method of successive approximations may be used for fitting curves to nonlinear relationships, only the method of least squares is described.

The decision to discuss the analytics and the curve-fitting methods in separate sections of the Memorandum was purely arbitrary. For many purposes, the user of this material will want to combine his readings in the first section with his readings in subsequent sections. The parallel nature of the presentations in each section was designed to facilitate this.

CONTENTS

TABLES

FIGURES

I. CURVE FITTING AND EMPIRICAL EQUATIONS

INTRODUCTION

An effective cost analysis capability cannot exist without the systematic collection and analysis of data on past, present, and projected programs. The analysis must result in the development of estimating relationships which can be used as a basis for estimating the resource impact of future proposals. These relationships typically relate resource requirements to the physical, performance, or operational characteristics of the system, and are, in essence, formal statements of the way one or more variables relate to each other. At times a simple factor such as cost per mile is all that is needed. Frequently, however, a more complex relationship is required such as the one between the weight and cost of an aircraft and the manhours required to maintain it. The necessary relationships are usually expressed as mathematical equations, as curves drawn on coordinate paper, or both. In either case, the methods of curve fitting are essential to the development process.

Just what do we mean by curve fitting? Suppose we plot a set of corresponding values of two variables on coordinate paper. The problem of curve fitting is that of finding the equation of a curve that passes through (or near) these points in the graph so as to indicate their general trend. An equation determined in this way is called an empirical equation between two variables, and the process of finding it is called curve fitting. While curve fitting as such deals primarily with relationships between two variables, certain of the

basic methods can be used to establish empirical equations among three or more variables. In cost analysis it is often useful to show the relationship between two variables in the form of a curve even when there is no mathematical expression possible; the methods of curve fitting can be used to establish such curves.

One of the interesting things about curve fitting is that there are so many different ways to do it, and a review of the literature on the subject would lead one to believe that there are as many methods as there are authors and problems. In fact, we may reasonably conclude that curve fitting is more of an art than a science. Fortunately, however, many of the methods are useful in solving only a limited number of unique problems, and for that reason are not of interest to us here. It is the intent of this Memorandum to explain curve fitting in a general sense and to present only those methods that experience has shown to be of general use to the cost analyst.

SOME BASIC ANALYTIC GEOMETRY

This section discusses the mathematical properties of some functional forms, the general shape of the curves portrayed by each, and the relationship between the shape of the curve, its location, and the values of the equation constants. Since our greatest concern at this point is to develop the equation describing a particular relationship, we will present the techniques for calculating the equation parameters, both constants and coefficients.

The different functional forms that will be treated are summarized below:

$y = a + bx,$	straight line,
$y = a + bx + cx,^2$	parabola,
$y = ab^x,$	exponential,
$y = ax^b,$	power function.

These suffice for most cost analysis problems with two variables. Although there are other forms that are frequently used, they generally can be related to those given above through appropriate scale transformations.

Straight Line

The straight line is certainly the simplest functional form to deal with. It is completely defined by knowing any two points on the line. The main feature of any straight line is the slope or tilt of the line. If the line rises reading from left to right as in Fig. 1, from points x_1y_1 to $P(x,y)$, the slope is said to be positive; if the line falls reading from left to right, the slope is said to be negative. The actual value of the slope (b) is the ratio of a change in y to a related change in x, that is, the ratio of the length of the vertical dashed line to the length of the horizontal dashed line in Fig. 1.

If we are given any point on a line, say $P(x,y)$ as in Fig. 1, and the slope (b), we would be able to deduce the equation of the line. This may be expressed symbolically as

$$b = \frac{(y - y_1)}{(x - x_1)},$$

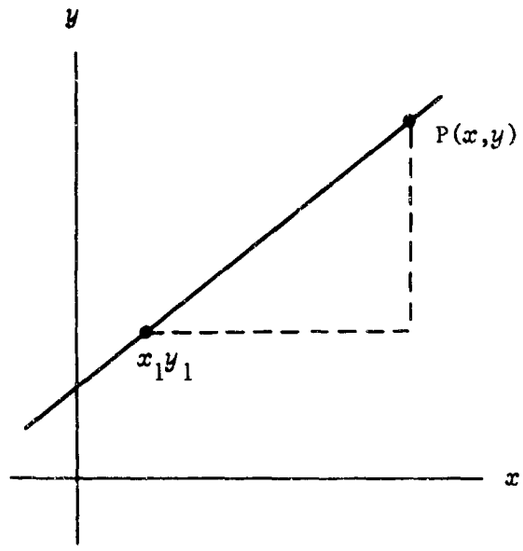


Fig. 1--Straight line with known slope drawn through point x_1y_1.

or

$$(y - y_1) = b(x - x_1),$$

which is known as the Point-Slope form of the equation of a straight line. If we are given one point and the slope, we would immediately be able to substitute in the above and have the equation of the line.

A slight modification of this results when, as in Fig. 2, the slope is not known directly, but two points on the line are given. Then the slope can be calculated as follows:

$$b = \frac{(y_2 - y_1)}{(x_2 - x_1)},$$

and this value may in turn be substituted into the Point-Slope formula.

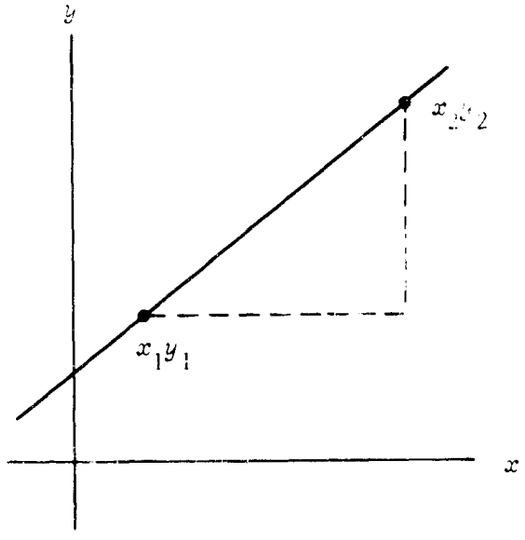


Fig. 2--Straight line drawn through two points, x_1y_1 and x_2y_2

The modified form of the Point-Slope formula for the equation of a straight line is

$$(y - y_1) = \frac{(y_2 - y_1)}{(x_2 - x_1)} (x - x_1).$$

This particular form is probably used more than any other in fitting straight lines.

There are other instances when, as in Fig. 3, the slope and the intercept are known. The intercept, more properly called the y intercept, is the point where the line crosses the y axis. This point is identified in Fig. 3 as $P(0, y)$. Because the coordinates of the intercept are as useful as the coordinates of any other point, we may use them to write the equation of the line. The Point-Slope formula for

a straight line is used and the result is

$$(y - a) = b(x - 0)$$

which simplifies to $y = a + bx$ where both the intercept (a) and the slope (b) are immediately recognizable. This is known as the Slope-Intercept formula for the equation of a straight line.

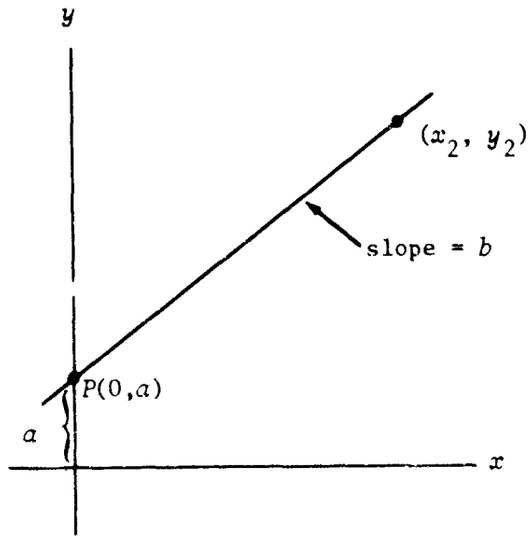


Fig. 3--Straight line with y intercepting at a and passing through point x_2y_2

The remaining case is where both the x and the y intercepts are known. As in Fig. 4, a is the value of y when x is equal to 0, and c is the value of x when y is equal to 0. Notice that, in this case, the line slopes downward from left to right so that we would expect the slope to be negative. Writing the equation for calculating the slope using the modified Point-Slope formula, we have

$$(y - a) = \frac{(a - 0)}{(0 - c)} (x - 0),$$

$$y = a - \frac{a}{c} x.$$

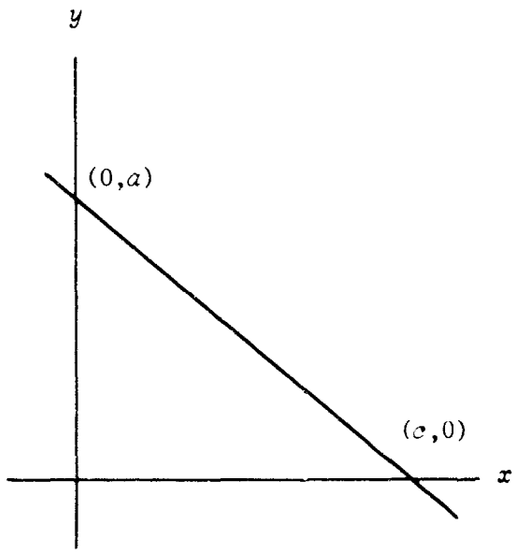


Fig. 4--Straight line showing x and y intercepts

Notice that the y intercept is equal to a and the slope is equal to a divided by c. This form of the formula for a straight line is called the Intercept form. Another form of this equation which results from a slight rearrangement is

$$\frac{x}{c} + \frac{y}{a} = 1.$$

In this form the coordinates of the two intercepts are immediately recognizable.

The next figure, Fig. 5, shows two special cases of the straight line, the line parallel to the x axis and the line parallel to the y

axis. In the first case the slope, m, is 0 and in the second it is infinite. The equations for these two lines are quite simple. In the first case where the line is parallel to the x axis, the equation is

$$y = a,$$

where a is the constant distance from the x axis. In the second case where the line is parallel to the y axis, the equation is

$$x = c,$$

where c is the constant distance of the line from the y axis.

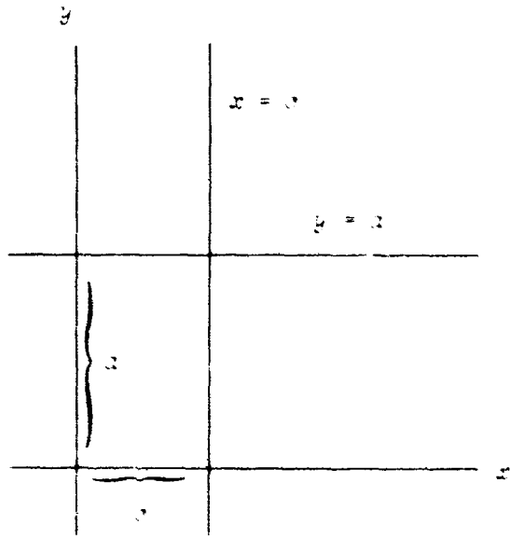


Fig. 5--Straight line parallel to x or y axis

Any of the formulas presented above may be used to write the equation of a straight line. The choice will depend on the particular kind of information available at the time, for there are times

when each is useful. The most generally useful, however, is the modified Point-Slope formula.

Parabola

The parabola is not as commonly used as the straight line, but has sufficient application to make it worthy of treatment here. The parabola is defined as the curve described by points equidistant from a fixed point and a fixed line. Figure 6 shows such a curve. The fixed point F is known as the focus of the parabola and the fixed line is as the directrix. Of course the equation of such a curve depends on its location with respect to the coordinate axes. For the moment, however, we will position the curve as shown in Fig. 7. The vertex

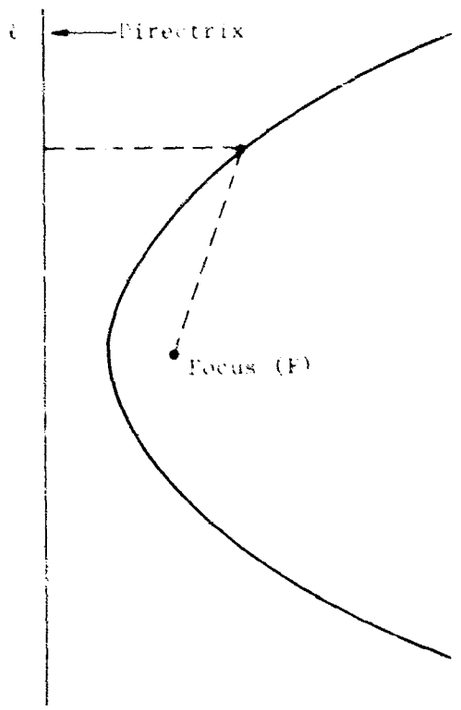


Fig. 6--A parabola

is at the origin of the coordinate axes and the line of symmetry of the curve is the x axis. Referring to the definition, we find that when the value of y is 0 (which is the case at the origin), the directrix is the same distance to the left of the origin as the focus is to the right. Further, assuming the distance from the directrix to the focus on the line of symmetry is p, the coordinates of the focus are by definition $(p/2,0)$, and similarly the equation of the directrix is $x = -p/2$.

Letting P be any point on the parabola and setting the distances FP and PL equal to each other, according to the definition, we can derive the equation for parabolas symmetrical to the x axis, the

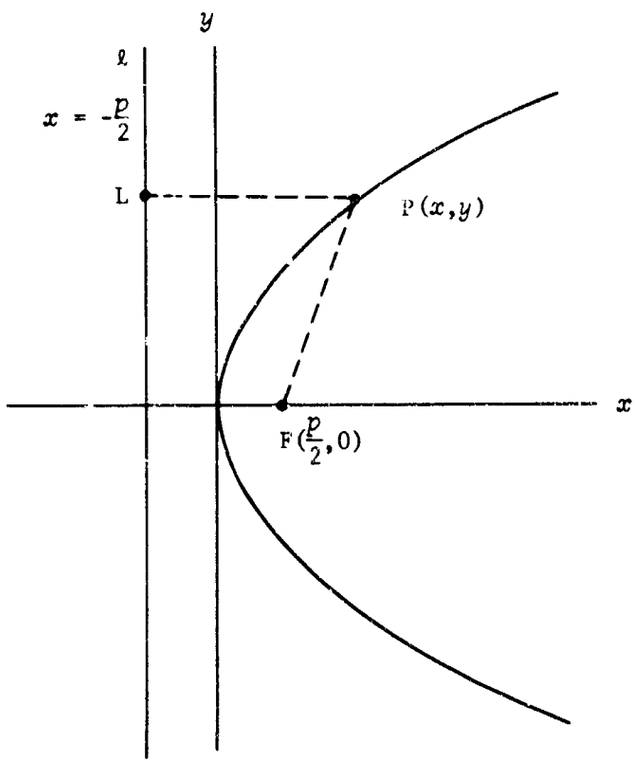


Fig. 7--Parabola with vertex at the origin, opening to the right

vertex at the origin, and opening outward to the right. Using the standard distance formula for determining the length of the two lines FP and PL, and setting one equal to the other,* we obtain

$$FP = \sqrt{(y - 0)^2 + (x - p/2)^2},$$

$$PL = \sqrt{(x + p/2)^2}.$$

$$\sqrt{(y - 0)^2 + (x - p/2)^2} = \sqrt{(x + p/2)^2}.$$

When both sides of this expression are squared and the result expanded, we have

$$y^2 + x^2 - px + \frac{p^2}{4} = x^2 + px + \frac{p^2}{4},$$

---

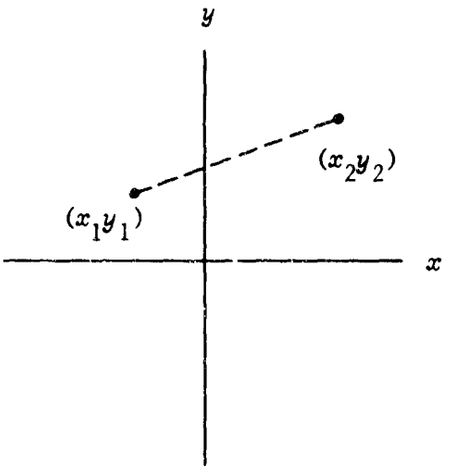
*The distance between any two points on rectangular coordinates may be calculated by using the following formula:

$$d = \sqrt{(y_1 - y_2)^2 + (x_1 - x_2)^2},$$

where d = the required distance,

(x_1, y_1) = the coordinates of the first point,

(x_2, y_2) = the coordinates of the second point.



which simplifies to

$$y^2 = 2px,$$

and is the equation of the parabola shown in Fig. 7. If the parabola were as pictured in Fig. 8, the equation could be obtained by using the same method:

$$FP = \sqrt{(y - p/2)^2 + (x)^2},$$

$$PL = \sqrt{(y + p/2)^2},$$

$$\sqrt{(y - p/2)^2 + (x)^2} = \sqrt{(y + p/2)^2},$$

$$y^2 - py + \frac{p^2}{4} + x^2 = y^2 + py + \frac{p^2}{4},$$

$$x^2 = 2py,$$

which is the equation for a parabola with its vertex at the origin, symmetrical to the y axis, and opening upward. Notice that the ninety-degree rotation, as was made between Fig. 7 and Fig. 8, caused the x and the y terms to be interchanged. Otherwise the two equations are identical.

When the vertex of the parabola is shifted away from the origin, as in Fig. 9, the equation will again be altered. To show how, we regard the problem as one of shifting the intersection of the axis of the coordinate system from the point (h,k) to the point $(0,0)$. To make the translation we set

$$x = x' + h \quad \text{or} \quad x' = x - h,$$

$$y = y' + k \quad \text{or} \quad y' = y - k,$$

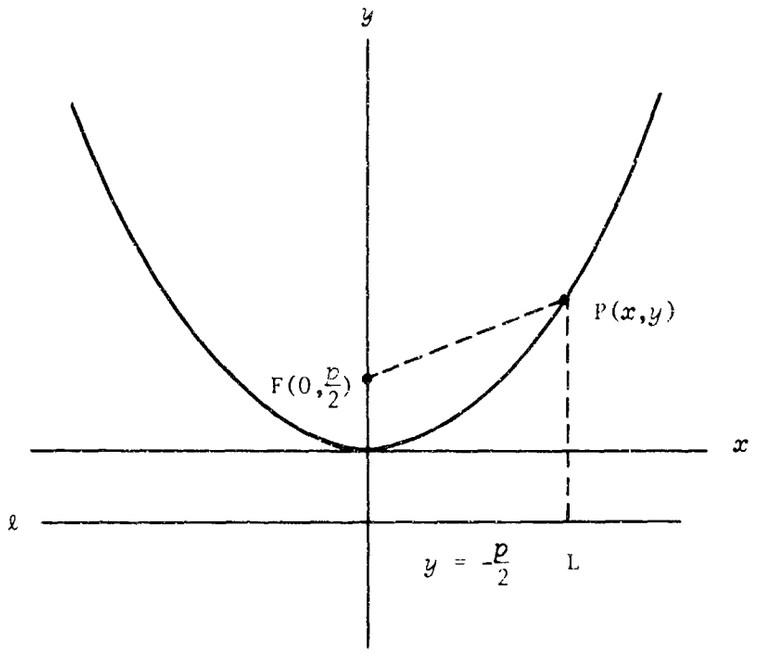


Fig. 8--Parabola with vertex at the origin, symmetrical to the y axis and opening upward

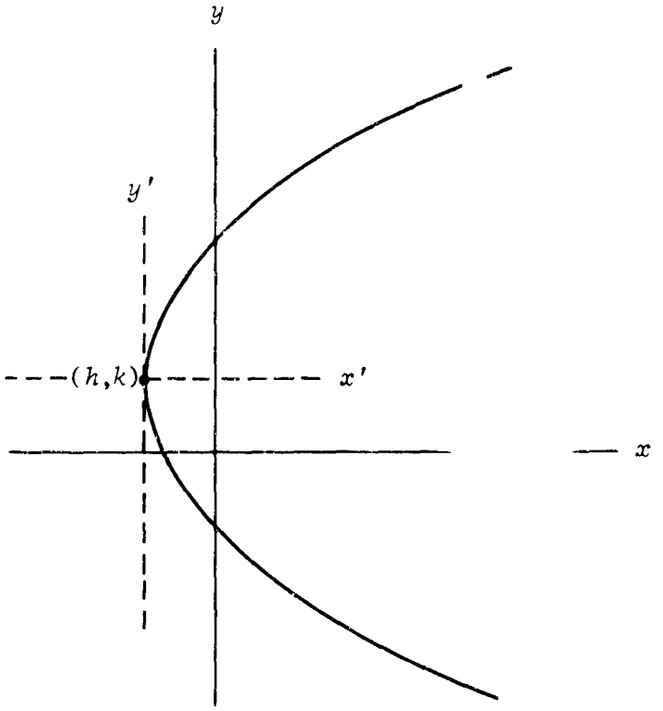


Fig. 9--Relationship between parabola and two sets of coordinate axes

where x and y refer to the original axes; x' and y' refer to the axes whose center coincides with the vertex of the parabola; and h and k are the coordinates of the origin of the x', y' axes measured from the x,y axes.

When we substitute $(x - h)$ for x' and $(y - k)$ for y' in the equation $y'^2 = 2px'$, we have

$$(y - k)^2 = 2p(x - h),$$

which when expanded yields

$$y^2 - 2ky + k^2 = 2px - 2ph,$$

$$y^2 - 2px - 2ky + 2ph + k^2 = 0.$$

Because in all cases h, k, and p will be constants, the equation may be written as follows:

$$y^2 + Dx + Ey + F = 0,$$

where $D = -2p$,

$E = -2k$,

$F = 2ph + k^2$.

This is the standard form of the equation for all parabolas symmetrical to a line parallel to the x axis.

If instead of $y'^2 = 2px'$ we had started with $x'^2 = 2py'$, we would arrive at the standard form of the equation as follows: Substituting $(y - k)$ and $(x - h)$ for y' and x' respectively gives us

$$(x - h)^2 = 2p(y - k),$$

which when expanded yields

$$x^2 - 2xh + h^2 - 2py + 2pk = 0.$$

After substituting as above, we have

$$x^2 + Dy + Ex' + F' = 0,$$

where $D = -2p$,

$E' = -2h$,

$F' = 2pk + h^2$.

This is the standard form of the equation for all parabolas symmetrical to a line parallel to the y axis.

If we take each of the two standard forms in turn, shift the terms and divide through appropriately, we arrive at the following equations:

$$\frac{1}{D}y^2 + \frac{E}{D}y + \frac{F}{D} = x,$$

$$\frac{1}{D}x^2 + \frac{E'}{D}x + \frac{F'}{D} = y.$$

As each of the coefficients in the above expressions is a constant, we can make further substitutions and obtain either

$$Ay^2 + By + C = x,$$

or

$$Ax^2 + Bx + C = y,$$

where $A = 1/D$,

$B = E/D$ or E'/D,

$C = F/D$ or F'/D.

These are the forms of the parabola that are most commonly used in curve fitting. Since there are three coefficients, or unknowns, at least three points must be known to define the curve. Given three points on a parabola, the equation may be obtained by using the coordinates of each of the points to obtain an equation of the above form and then solving the three equations simultaneously for A, B, and C.

To illustrate, we are given the three points (0,2), (3,4), and (4,12). Plotting these points as in Fig. 10 leads us to believe a parabola opening upward and symmetrical to a line parallel to the y axis would be the correct form to fit. The standard form of the equation for this type of parabola is

$$y = Ax^2 + Bx + C.$$

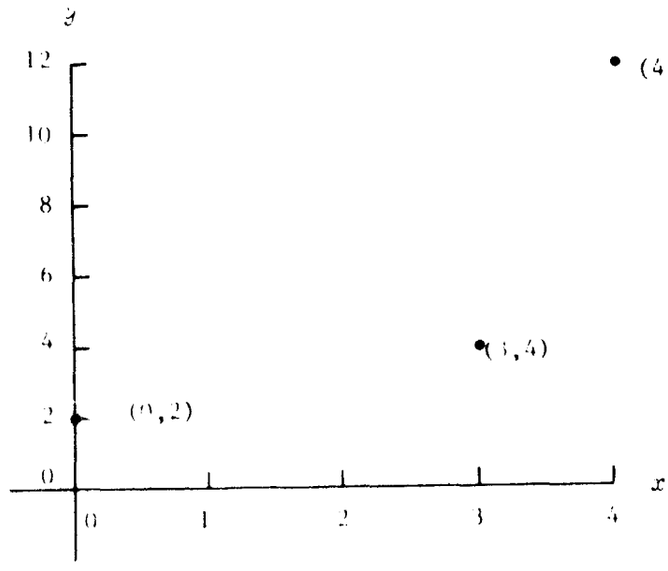


Fig. 10--Points used to illustrate fitting the parabola

Substituting each of the three points in this expression allows us to write the following three equations. Notice that the x coordinate must be squared to make certain of the substitutions:

$$2 = 0A + 0B + C,$$

$$4 = 9A + 3B + C,$$

$$12 = 16A + 4B + C.$$

It is obvious from the first equation that C is equal to 2. Using this knowledge to adjust the two remaining equations will reduce the problem significantly. In this case, the two remaining equations are

$$2 = 9A + 3B,$$

$$10 = 16A + 4B.$$

There are a number of ways to solve simultaneous equations. Probably the simplest for only two equations is the determinant method. As the number of variables and the number of equations get larger, however, other methods are preferred. In fact, when four or more equations are involved, it is probably best to look for computer programs to do the job. The determinant method is particularly well adapted to the desk calculator, but not particularly well suited for illustrative purposes. Here we will divide by the leading coefficients and eliminate variables by subtraction.

Dividing the first equation by 9, the second by 16, and subtracting the first equation from the second, A is eliminated. These steps follow:

$$A + \left(\frac{3}{9}\right)B = \frac{2}{9},$$

and

$$A + \left(\frac{4}{16}\right)B = \frac{10}{16}.$$

Subtracting the first from the second, we get

$$\left(\frac{1}{4} - \frac{1}{3}\right)B = \left(\frac{5}{8} - \frac{2}{9}\right).$$

The necessary simplifications and other arithmetic having been performed,

$$B = -\frac{29}{6}.$$

We next substitute the value of B into the first of the two variable equations and calculate A as follows:

$$A + \left(\frac{3}{9}\right)\left(-\frac{29}{6}\right) = \left(\frac{2}{9}\right),$$

$$A = \frac{2}{9} + \frac{87}{54},$$

$$A = \frac{108 + 783}{486},$$

$$A = \frac{11}{6}.$$

The required coefficients are now seen to be

$$A = \frac{11}{6},$$

$$B = -\frac{29}{6},$$

$$C = 2.$$

It is usually good practice to substitute all of these coefficients into one of the original equations to test the correctness of the arithmetic. Substituting in the second equation we have

$$4 = 9\left(\frac{11}{6}\right) + 3\left(-\frac{29}{6}\right) + 2.$$

The required arithmetic shows us that the values of the coefficients calculated are in fact correct. The equation that we have been looking for is therefore

$$y = \left(\frac{11}{6}\right)x^2 + \left(-\frac{29}{6}\right)x + 2.$$

Solving this equation for y, given a range of values of x and plotting them, allows us to draw the curve shown in Fig. 11. Contrary to our expectation, this form of parabola is not a good representation of the relationship implied by the three points. This example illustrates

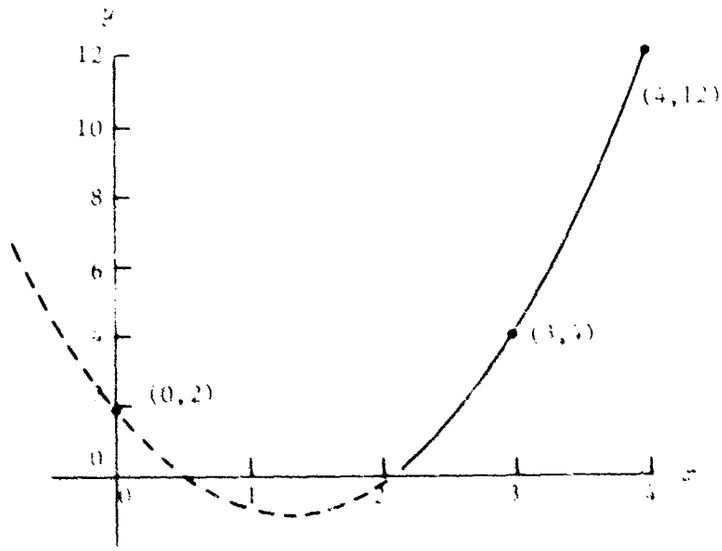


Fig. 11--Parabola passing through three points

an inherent difficulty associated with using the parabola. If we had not examined the characteristics of this curve over the relevant values of x, the fact that y is negative between $x = 1\frac{1}{2}$ and $x = 2$ would not have been noticed and could have led to absurd cost estimates.

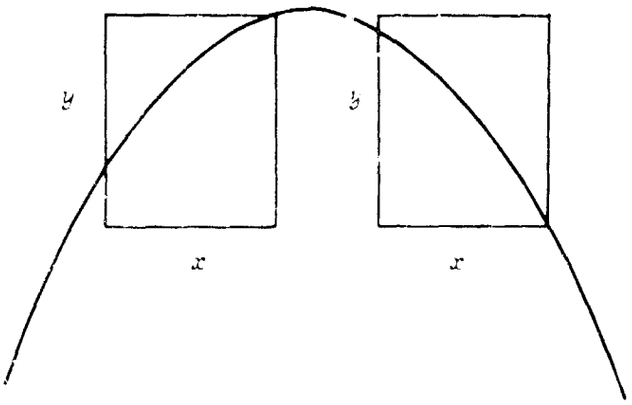
In curve fitting we are concerned primarily with the best representation of the data at hand. In cost estimating we are typically concerned with extrapolation beyond the range of the existing data. When we choose a parabola to represent a relationship between two sets of data, we generally use only a limited segment of the entire curve. Figure 12 illustrates how this fact can lead to trouble The boxed-in segments of the curve show the part of the curve used to describe the data. Examination of the curves outside the limits of the various boxes shown in Figs. 12a, 12b, and 12c indicates the kind of trouble one can get into by using this type of curve for making extrapolations.

There are times when the best parabolic function to represent a set of data is of the form

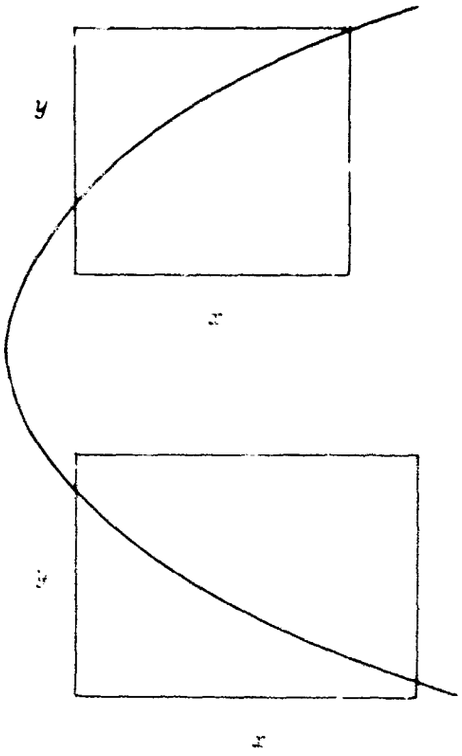
$$ax^2 + bx + c = y$$

Since it is conventional for y to be the dependent variable, this equation causes some difficulty. One way this difficulty can be overcome, after fitting the curve, is to solve the resultant equation for x using the quadratic formula. First the equation must be rewritten as follows:

$$ax^2 + bx + (c - y) = 0.$$

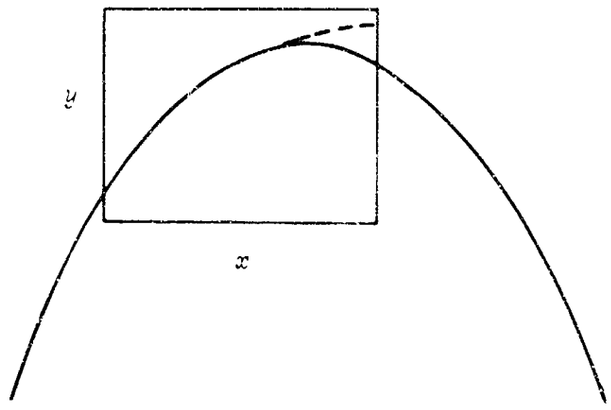


(12a) Parabola opening downward



(12b) Parabola opening to the right

Fig. 12--Implications of the use of parabolas for extrapolation



(12c) Another parabola opening downward

Then, using the quadratic formula,

$$y = \frac{-B \pm \sqrt{B^2 - 4(A)(C-x)}}{2A} .$$

Use of this formula will probably result in two solutions because the square root of a number can be either positive or negative. Each equation must be evaluated to determine which is appropriate.

Exponential

The general form of the exponential equation is $y = ab^x$.* Graphs of two exponential equations that differ from each other only with respect to the value of b are shown in Fig. 15. In each case a has been set equal to 1. As will be shown, only the level of the exponential is affected by the value of a.

A graph similar to that shown in Fig. 13a results wherever b is greater than 1, and a graph similar to that in Fig. 13b if b is between

*In this text, the function with the independent variable x as the exponent is called the exponential, while the function $y = ax^b$, where the exponent is a constant, is called the power function.

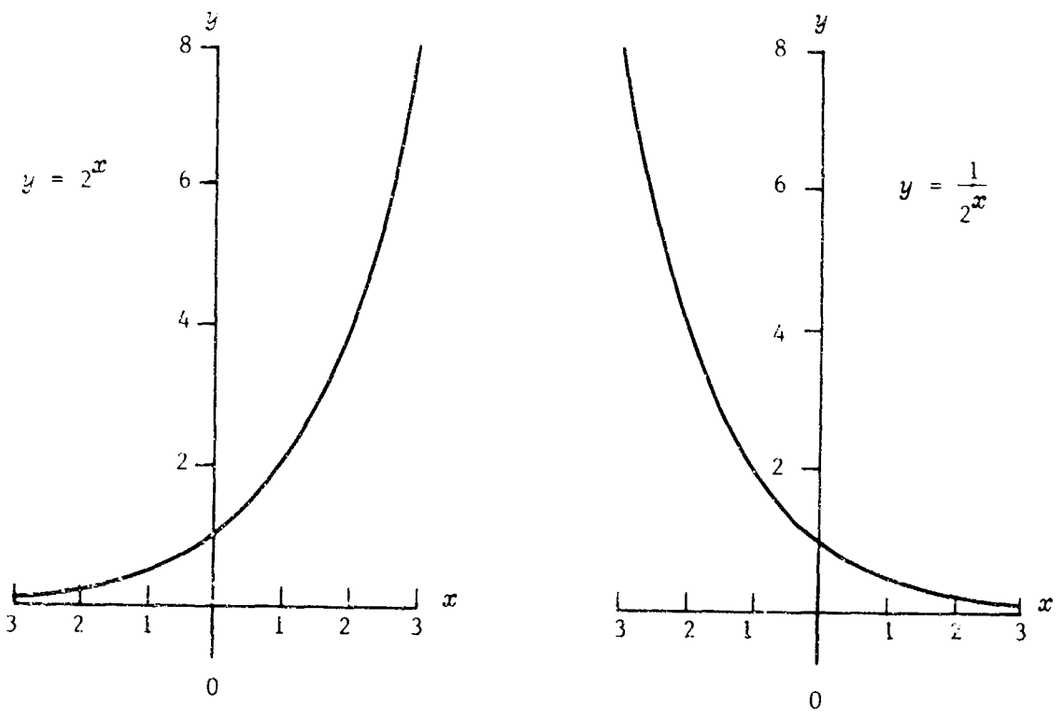
1 and 0. If b is equal to 1 the exponential equation becomes

$$y = 1,$$

for 1 raised to any power is equal to 1. If b is 0 there is no equation, for 0 raised to any power is 0, and consequently

$$y = 0.$$

When b is negative (less than 0), the exponential is discontinuous and, for that reason, of no value to us for curve fitting.



(13a) Exponential with $b > 1$ (13b) Exponential with $0 < b < 1$

Fig. 13--Negative and positive exponential curves

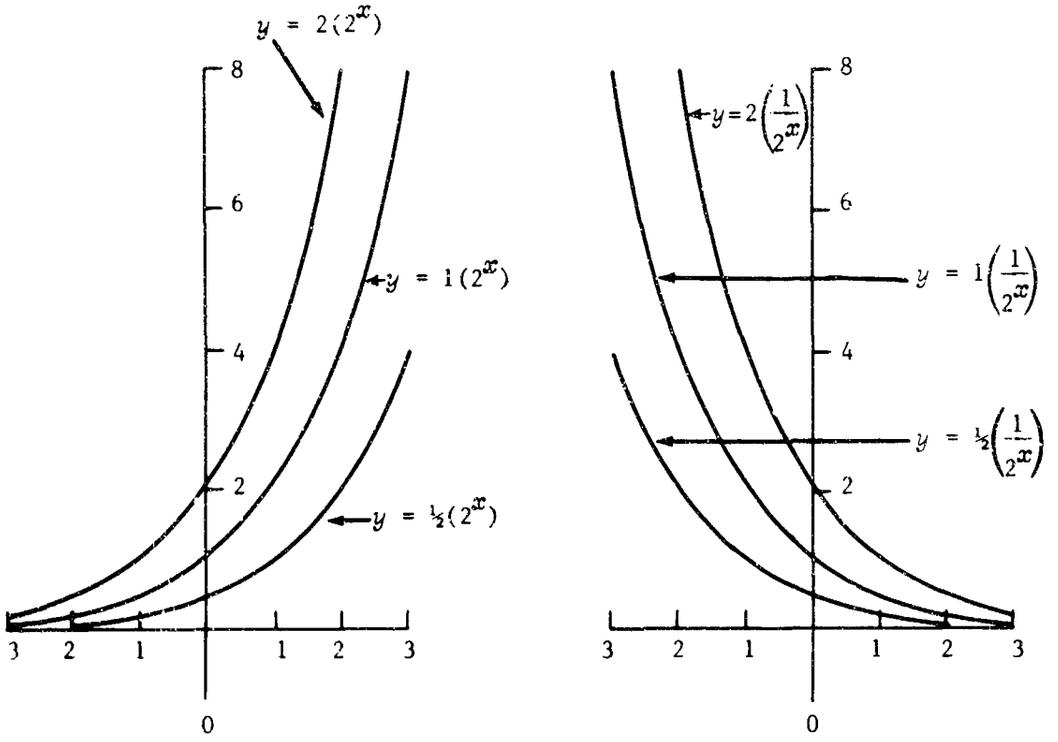
For our purposes, the fact that the exponential curve rises from left to right when b is greater than 1 and from right to left when b

is between 1 and 0 is the relevant characteristic. Also notice that both curves pass through the point $y = 1$, $x = 0$.

The influence of a is illustrated in Fig. 14 Larger values tend to raise the curve while lower values cause a downward shift. When x is equal to 0, $b^0 = 1$, and the exponential becomes

$$y = a.$$

Consequently, a may be thought of as the y intercept.



(14a) Exponential with $b > 1$ (14b) Exponential with $0 < b < 1$

Fig. 14--The effect of the values assigned to a on the level of the exponential

Since facility with the exponential requires an understanding of exponents and logarithms, we will digress temporarily to review these

topics. The system of exponents is based entirely on five basic laws and four definitions. The first definition states that the expression a^n, where n is an exponent and a is greater than 0, is the product of a multiplied by itself n times:

$$a^2 = a \times a,$$

$$a^3 = a \times a \times a,$$

etc.

The first law of exponents states that the product of a^m and a^n is a^{m+n} which incidentally follows directly from the initial definition. To illustrate:

$$a^2 \times a^3 = a^{2+3} = a^5,$$

$$a^2 = a \times a, \quad a^3 = a \times a \times a,$$

$$(a \times a)(a \times a \times a) = (a \times a \times a \times a \times a) = a^5.$$

Each of the other laws can be similarly derived and it would be a worthwhile exercise for the reader to do so. All five laws are summarized below:

I. $a^m \times a^n = a^{m+n}$.

II. $a^m / a^n = a^{m-n}$.

III. $(a^m)^n = a^{m \times n}$.

IV. $(ab)^n = a^n \times b^n$.

V. $(a/b)^n = a^n / b^n$.

Three additional definitions complete the system; a^0 is defined as 1, a^{-n} is defined as $1/a^n$,* and $a^{1/n}$ is defined as the nth root of a. The root is positive if a is positive, and negative if a is negative and n is odd. This system not only gives meaning to the expression a^x when a is greater than 0 and x is any rational number, but also provides the inputs essential to a discussion of logarithms.

The logarithm of a number is the power to which a base number must be raised to equal the original number; it can be more conveniently expressed as

$$y = a^x,$$

where x is the logarithm of y to the base a. In the language of logarithms we would write

$$x = \log_a y.$$

The logarithm x is also an exponent. From this and our earlier discussion of exponents, we conclude, and rightly so, that any rational number greater than 0 can be the base of a system of logarithms. In actual practice, however, 10 and the constant e are most commonly used.** When the base is 10, the logarithms are called *common* (logs)

---

*It should also be pointed out that such definitions seem logical from the law of division. That is,

$$1 = \frac{a^n}{a^n} = a^{n-n} = a^0 \quad \text{and} \quad \frac{1}{a^n} = \frac{a^0}{a^n} = a^{0-n} = a^{-n}.$$

**The constant e is the limit of the expression $(1 + v)^{1/v}$ as v approaches 0; the limit is equal to 2.7183 to five significant figures. It is one of the most important limits in calculus.

logarithms, and when the base is e they are called *natural* (lns), or Napierian logarithms. We shall follow the general practice of using the abbreviation log where 10 is the base and ln where e is the base. Tables of each are readily available.

Any rational number greater than 0 can be expressed in terms of its logarithm and consequently in terms of 10 or e. Expressing a relationship in terms of e leads to simplification both of form and of required computations. Suppose, for example, we have a number, y, which we wish to express in terms of e. We would only have to find ln y in a table of natural logarithms to write

$$\ln y = x,$$

or in exponential form,

$$y = e^{x}.$$

Figure 15(a) shows us that these two equations have exactly the same graph as do the equations

$$\ln x = y,$$

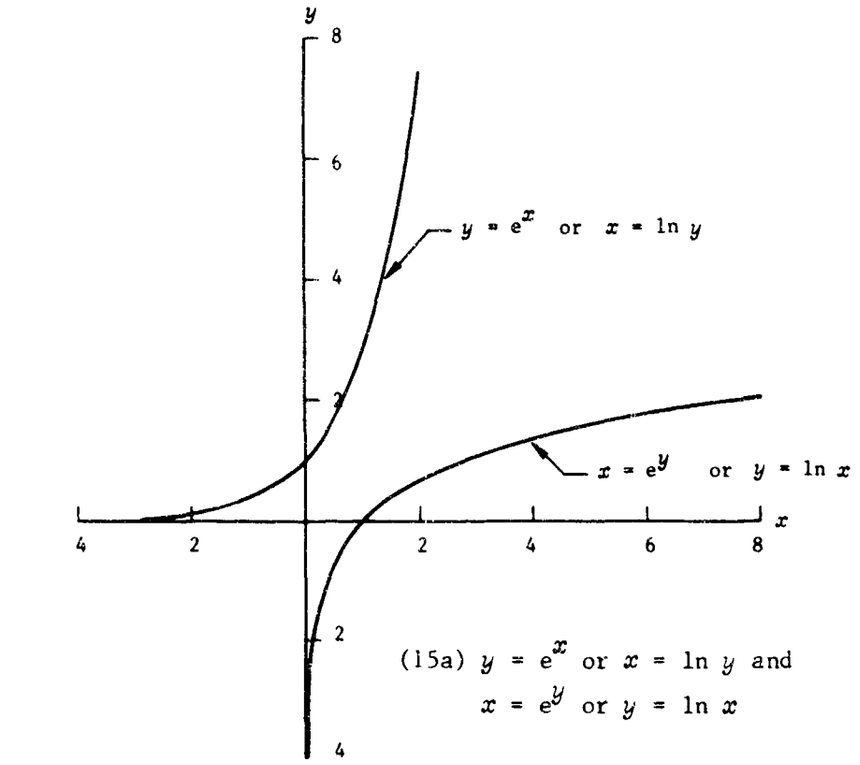
and

$$x = e^{y}.$$

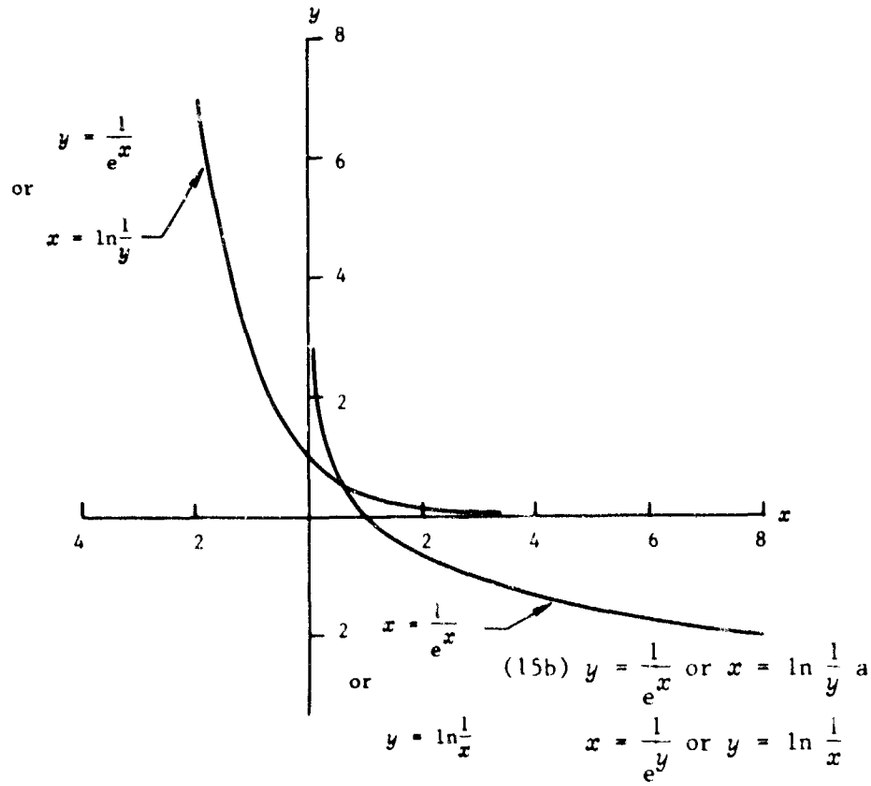
Fig. 15(b) provides similar information for reciprocal relationships.

Interchanging the x and y terms does, however, cause an exchange of coordinate axes.

To express the exponential $y = 16.5^{x}$ in terms of e we treat the number 16.5 as e^{k} and write



(15a) $y = e^x$ or $x = \ln y$ and
$x = e^y$ or $y = \ln x$



(15b) $y = \frac{1}{e^x}$ or $x = \ln\frac{1}{y}$ and
$x = \frac{1}{e^y}$ or $y = \ln\frac{1}{x}$

Fig. 15--Curves illustrating the relationship between $y = e^x$, $x = e^y$, $y = 1/e^x$, and $x = 1/e^y$ and their logarithmic transformations

$$16.5 = e^{k}$$

and

$$\ln 16.5 = k.$$

From a table of natural logarithms we find that ln 16.5 is approximately 2.83 and we write either

$$\ln 16.5 = 2.83,$$

or

$$16.5 = e^{2.83}.$$

Substituting in the original exponential equation and applying the third law of exponents we obtain

$$y = (e^{2.83})^{x},$$

and

$$y = e^{2.83x}.$$

When the exponential is expressed in terms of e, the slope of the curve at any point is equal to the value of the expression at that point. When the exponential is not expressed as a function of e, the slope is proportional to, but not equal to, the value of the expression at the point; i.e., slope = ky.

For example, Fig. 16 shows the graph of the expression $y = 2(2^{x})$ which is an exponential of the form $y = b^{x}$. Since this is not written in terms of e, we would expect the slope at any point to be proportional to the value of the expression at that point. We can check this--at least approximately--by estimating the slope of the curve

at two points and comparing the results with the value of the function at those points. To do this we establish the equation

$$S_x = ky_x,$$

where S_x = the estimated slope at point x,

y_x = the value of the function $y = 2(2^x)$ at point x,

k = the constant of proportionality.

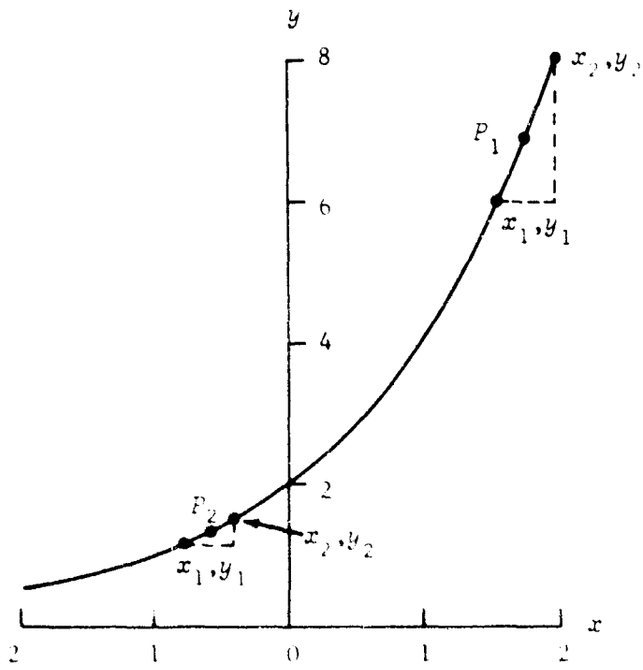


Fig. 16--Estimating the slope of the expression $y = 2(2)^x$ at points P_1 and P_2

Two points, P_1 and P_2, have been selected, one at either end of the curve. It is obvious that the slopes at these two points are different. Let us assume that the curve extended an equal distance from each P in either direction (shown on the graph as the hypotenuse of the indicated traingles) is a straight line. Remembering the discussion of

the straight line, we see that, having made this assumption, the coordinates of the vertexes of the triangles provide sufficient information to estimate the slope of the curve. The coordinates of the vertexes of the upper triangle are

$$x_2 = 2.00, \quad y_2 = 8.00,$$

$$x_1 = 1.60, \quad y_1 = 6.063.^{\star}$$

The formula for calculating the slope of a line given two points is

$$S \cong \frac{y_2 - y_1}{x_2 - x_1},$$

and on substitution

$$S \cong \frac{8.000 - 6.063}{2.00 - 1.60},$$

$$S \cong 4.843.$$

The x coordinate of point P_1 is the average of x_1 and x_2 (1.80). Where this point, x_a, is substituted in the expression $y = 2(2^{x_a})$, the resulting value of y is 6.964. Returning to our proportionality statement

$$S_x = \frac{y}{x},$$

we substitute appropriately and get

$$4.843 = \frac{6.964}{x},$$

$$x = 0.695,$$

---

*The y coordinates were obtained by substituting the x coordinates in the expression $y = 2(2^x)$; values to three decimal places were obtained by solving, a procedure that improves the agreement between the estimated slopes.

indicating that the slope of the curve can be evaluated at any point by multiplying the value of the function at that point by 0.695. To check this we use point P_2 in exactly the way we did P_1 and derive another estimate of the slope and the value of k.

For the lower triangle

$$x_2 = -0.4, \qquad y_2 = 1.516,$$

$$x_1 = -0.8, \qquad y_1 = 1.149;$$

therefore the slope

$$S \cong \frac{1.516 - 1.149}{-0.4 + 0.8},$$

$$S \cong 0.918.$$

The value of the function at $x_a(-0.6)$ is 1.320; therefore

$$0.918 = 1.320k,$$

$$k = 0.695;$$

and we are satisfied that the required constant of proportionality does exist. The value of the expression at each of the two points was calculated to four significant figures using the expression itself. Values were not read from the curve. It is also interesting to note that for the illustrative expression of the form $y = ab^x$, the value of b was 2. Further, the natural logarithm of 2 is equal to 0.6931 which is quite close to the value of k estimated above. The fact that our results were no closer to the theoretical value of k is due largely to the assumption that the curve was linear over the relevant range.

The constant of proportionality can be proved analytically to be exactly equal to the natural logarithm of b. Further, in an exponential of the form $y = e^x$, b is equal to e and the natural logarithm of e is equal to 1. Thus for these kinds of exponential expressions it is obvious that k must also be equal to 1.

Let us turn now from the digression to our main discussion. It has probably already been noticed that the exponential expressed in logarithmic form is a straight line function: i.e., the expression $y = ab^x$ is equivalent to the expression $\ln y = \ln a + \ln bx$; notice $\ln a$ and $\ln b$ are constants. This fact greatly facilitates fitting with exponential expressions.

If we are given the exponential $y = e^{a+bx}$ and we want to put it into logarithmic form, we take the logs of each side. Although the logarithm of y presents no problem, the logarithmic form of e^{a+bx} may appear to. When we remember the laws of exponents and the fact that logarithms are in fact exponents, we find that $a + bx$ is the logarithm of y to the base e, and as such becomes the natural form of the right-hand side of the equation

$$\ln y = a + bx.$$

The exponential expression is therefore linear when stated in terms of the logs of one of its members. To illustrate this, take the following expression

$$\ln y = a + bx.$$

We can recognize this as a semi-log straight line. If we were to

convert to exponential form we would have

$$y = e^{px+2}.$$

If we also examine the equation

$$y = x^{x+2},$$

we find that this is another form of the exponential which can be converted to log form as follows: First, we divide both sides of the equation by x which gives us

$$y/x = x^{x+2}.$$

We can also express x as a power of e. We look up the natural log of x, which, for lack of a better name, we will call p. We can now write

$$y/x = (e^{p})^{x+2},$$

or, again according to the third law of exponents,

$$y/x = e^{xp+2p}.$$

The expression may be further simplified by letting $2p$ be represented by the constant, c. This produces

$$y/x = e^{px+c}.$$

Now converting to logarithmic form we have

$$\ln (y/x) = px + c,$$

which once again is a linear expression when the quotient of y/a is given in terms of logarithms.

The equation of an exponential passing through two points may be determined quite simply. We need only set up the required functional form in terms of logarithms, then proceed with a straight line. The following example illustrates this method.

Given the two points, 1,7 and 4,1 on x,y coordinates, we choose an expression of the form $y = ab^x$ as the appropriate general exponential to fit. The next step is to restate this expression in terms of logarithms as follows:

$$\ln y = \ln a + x \ln b.$$

With the coordinates of the two points, (x_1,y_1) and (x_2,y_2) we may write the two equations:

$$\ln y_1 = \ln a + x_1 \ln b,$$

$$\ln y_2 = \ln a + x_2 \ln b.$$

Taking the logarithms of y_1 and y_2 and substituting the logs of the ys and the xs in the above equations results in two equations with two unknowns that may be solved simultaneously:

$$1.9459 = \ln a + 1 \ln b,$$

$$0.0000 = \ln a + 4 \ln b.$$

Subtracting the second equation from the first leaves

$$1.9459 = -3 \ln b,$$

$$\ln b = -0.6486.$$

This value can in turn be substituted into the first equation above with the result that

$$1.9459 = \ln a - 0.6486,$$

$$\ln a = 2.5945,$$

and with this we can write the required expression as follows:

$$\ln y = 2.59 - 0.649x.$$

The expression has been evaluated, and the results plotted in Fig. 17 pass exactly through the two points as required. We can simplify the expression by converting it to exponential form. To do this we must have the numbers represented by $\ln a$ and $\ln b$. Looking in a table of natural logarithms we find that

$$\ln a = 2.59 = \ln 13.4,$$

$$\ln b = -0.649 = \ln 0.523,$$

and we may write

$$y = 13.4(0.523)^x.$$

Notice that 13.4, the value of a in this expression, is in fact the y intercept. We can simplify still further by converting the expression to one in terms of e. To do this we think of 13.4 as e^r and 0.523 as e^s; thus

$$13.4 = e^r,$$

$$0.523 = e^s,$$

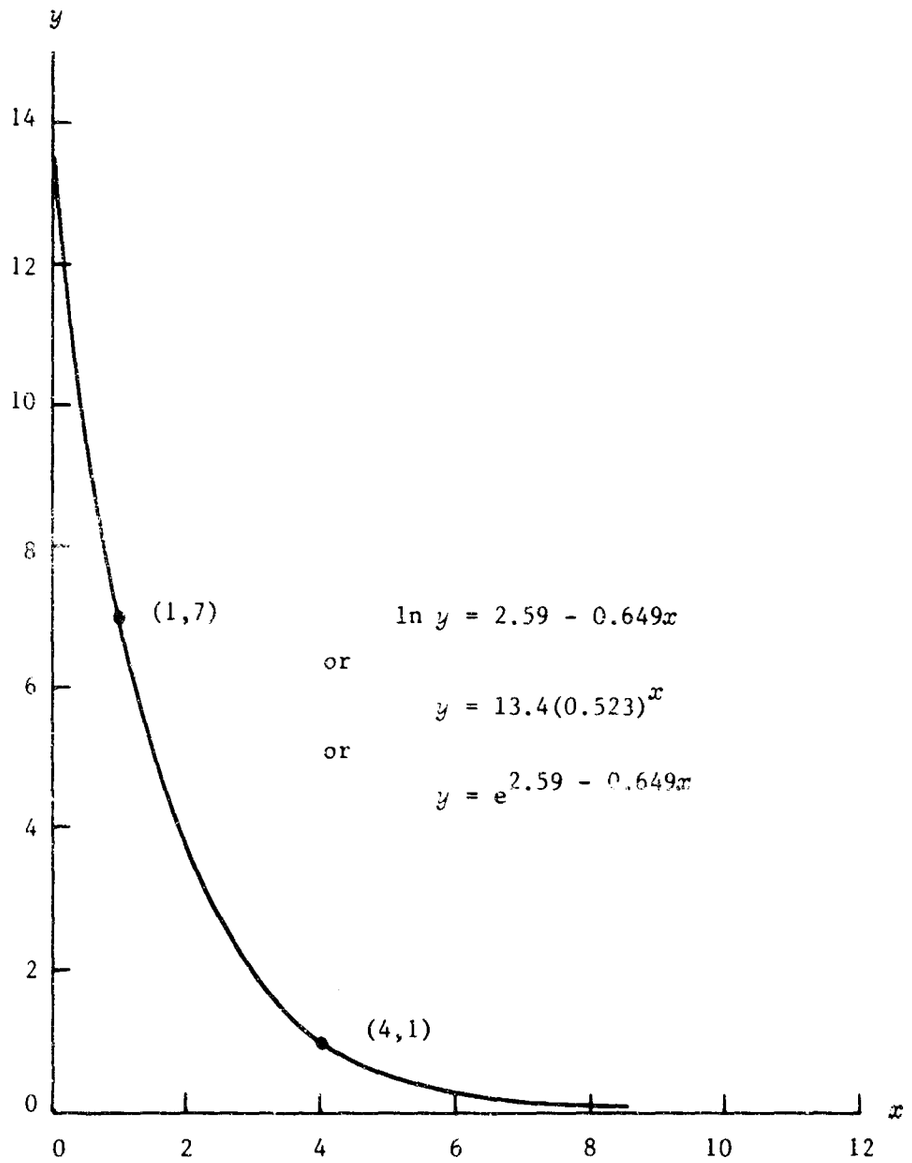


Fig. 17--Exponential fitted through two points

and we find the value of both r and s by again consulting a table of natural logarithms. Another way to write the last two expressions is

$$r = \ln 13.4$$

and

$$s = \ln 0.523.$$

We really did not need to use the table again because the lns of these values are already available from our calculations above:

$$s = \ln 0.523 = -0.649,$$

$$r = \ln 13.4 = 2.59.$$

Now if we substitute e^r and e^s in the equation $y = 13.4(0.523)^x$, we have

$$y = (e^{2.5})(e^{-0.649})^x.$$

Using the first and third laws of exponents we convert this expression to

$$y = e^{(2.59 - 0.649x)}.$$

Rewriting this expression in logarithmic form results in

$$\ln y = 2.59 - 0.649x,$$

that is, the same expression that we had initially.

Power Function

The power function is one of the most commonly used mathematical expressions in cost analysis because in many cases it adequately describes the phenomenon of decreasing costs of successive units of production. The general equation of this function is

$$y = ax^b.$$

To avoid confusing the power function with the exponential, which looks somewhat similar, we must observe the placement of the variable,

x. In the power function, the variable x is raised to the power b, while in the exponential, a constant is raised to the variable power x as below:

$$y = b^x.$$

The characteristics of the power function can best be illustrated by initially setting a equal to 1, because a affects primarily the level of the curve and has little influence on its shape. Having done this we are left with the equation

$$y = x^b.$$

For certain values of b, the function is not continuous for negative values of x. Therefore, we will restrict the variable x to values greater than 0; the exponent b can assume any value, positive, negative, or 0. However, when b is 0, the equation becomes

$$y = 1,$$

for any value raised to the 0 power is equal to 1. When b is positive and varies from 0, the family of curves shown in Fig. 18 results.

The smallest value assigned to b in Fig. 18 is 0.2. Had 0.0 been used, the result would have been a straight line parallel to the x axis and passing through $y = 1$ as above. When b is between 0 and 1 the curves generated are concave downward. When $b = 1$ a straight line ($y = x$) results, because any number raised to the first power is the number itself. As values greater than 1 are assigned to b, the curves become concave upwards. The curves pass through the point $x = 1$, $y = 1$ for all values of b, because 1 raised to any power always equals 1.

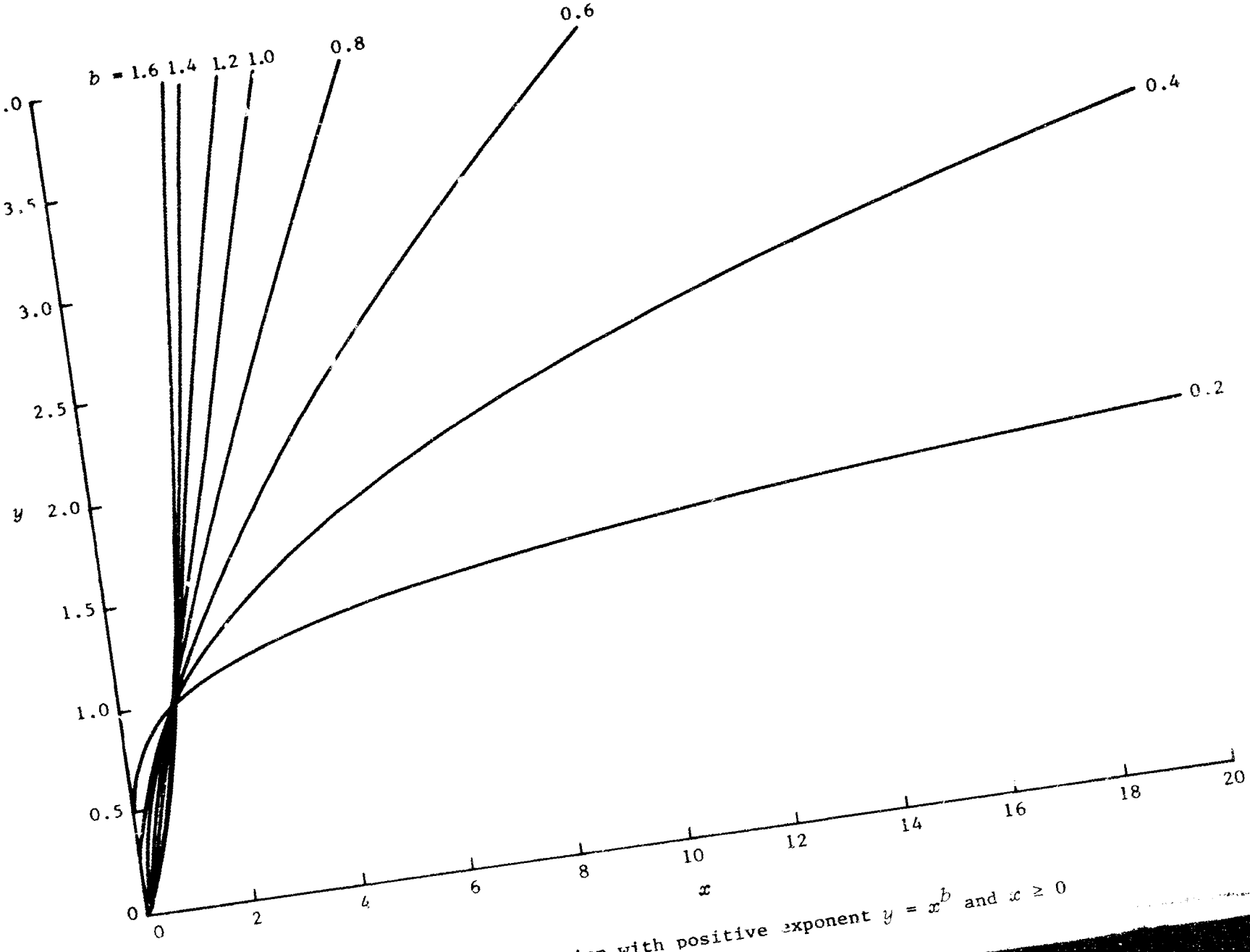


Fig. 18--Power function with positive exponent $y = x^b$ and $x \geq 0$

When x is greater than 1 the curves rotate upwards as the value of b increases. When x is between 1 and 0, however, the situation is not the same, as shown by Fig. 19. At the upper end all of the curves go through the point $x = 1$, $y = 1$ as before, but at the lower end they all tend towards the point $x = 0$, $y = 0$. When b is made smaller the curves become higher, the reverse of what happened when x was greater than 1. When b is greater than 1 the curves are concave upwards; when b is less than 1 the curves are concave downwards. As before, when b is equal to 1 the curve is the straight line $y = x$.*

When the exponent b is negative, the family of curves shown in Fig. 20 is generated. Regardless of the value of b when it is negative, the curves are concave upwards. As before, however, each of the curves passes through the point $x = 1$, $y = 1$, and for values of x greater than 1 the curves with the lower values of b lie above those with higher values of b. When x is less than 1, however, the reverse is true. When b is equal to -1 we do not have a straight line, as was the case when b was equal to +1; in this case the resulting equation is

$$y = \frac{1}{x},$$

which is a reciprocal or a form of hyperbola.

Figure 21 illustrates the effect of including a in the equation. When a is increased from 1 the curves shift upwards by direct multiplication. When a is reduced from 1 to 0 the curves shift similarly but in a downward direction.

---

*This is true because a decimal raised to a power greater than 1 gives a smaller number. Also, a decimal raised to a positive power less than 1 gives a larger number than itself.

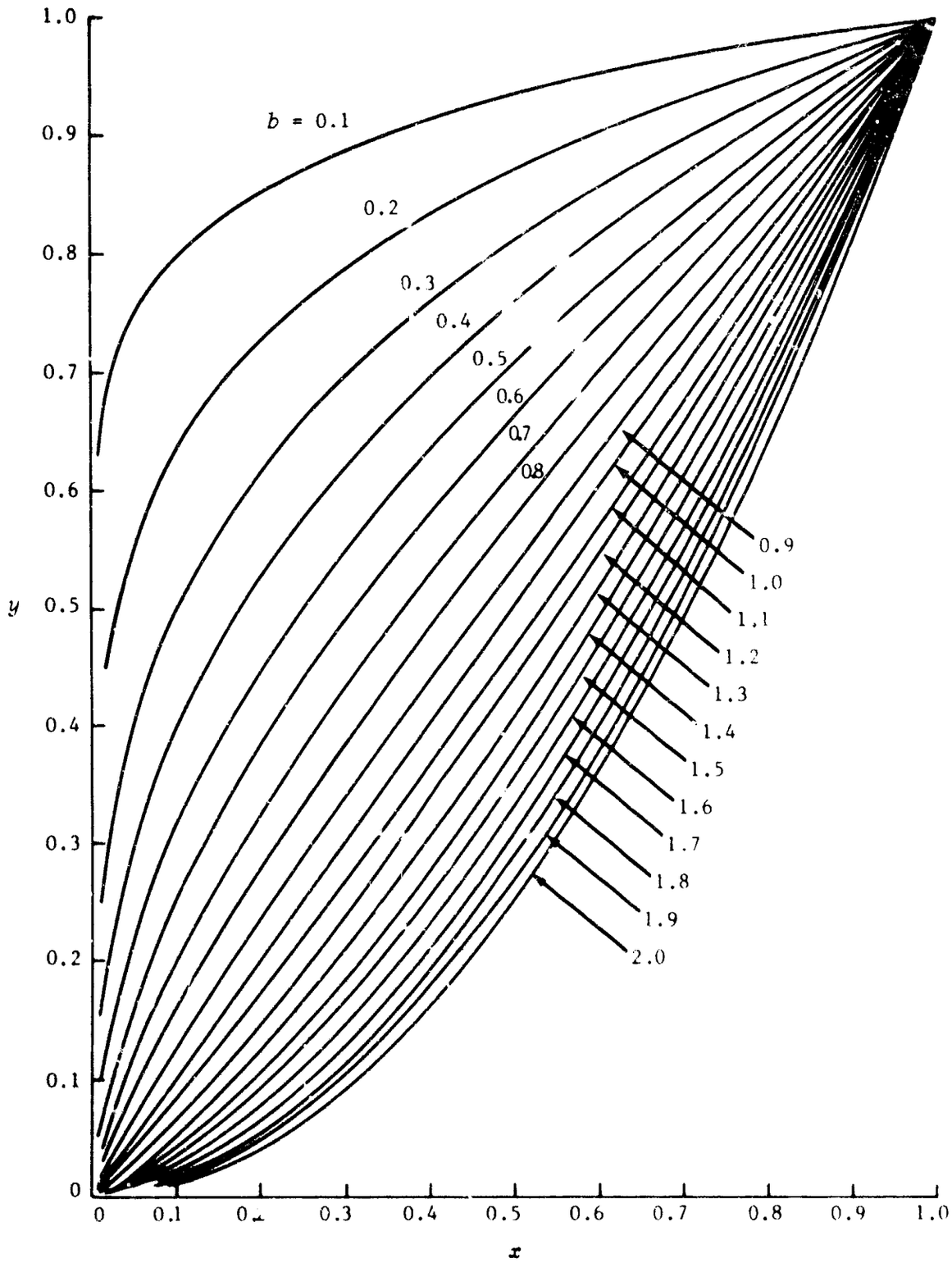


Fig. 1--Power function $y = x^b$, with positive exponents and $0 < x \leq 1$

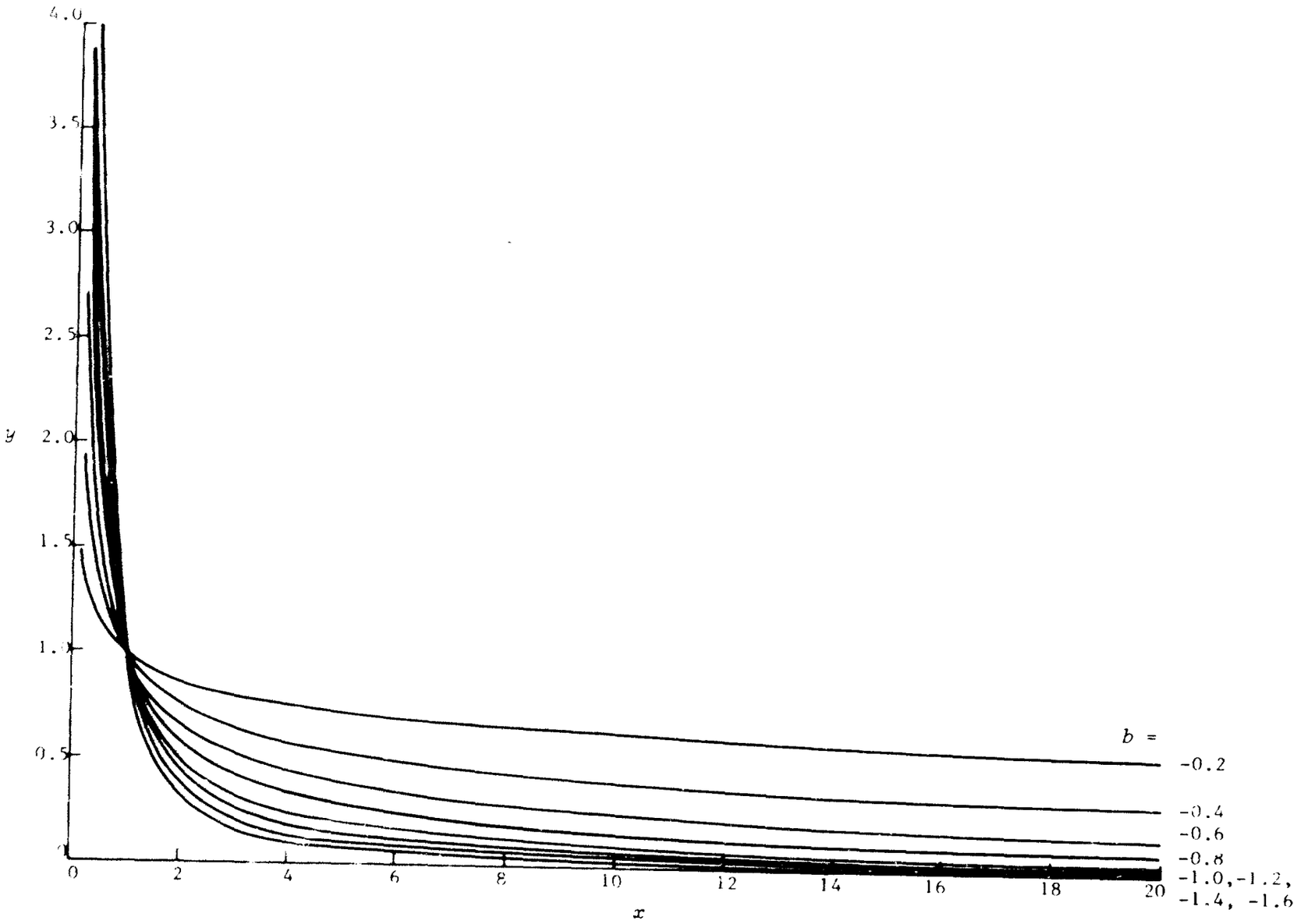


Fig. 20--Power function $y = x^b$, with negative exponent and $x > 0$

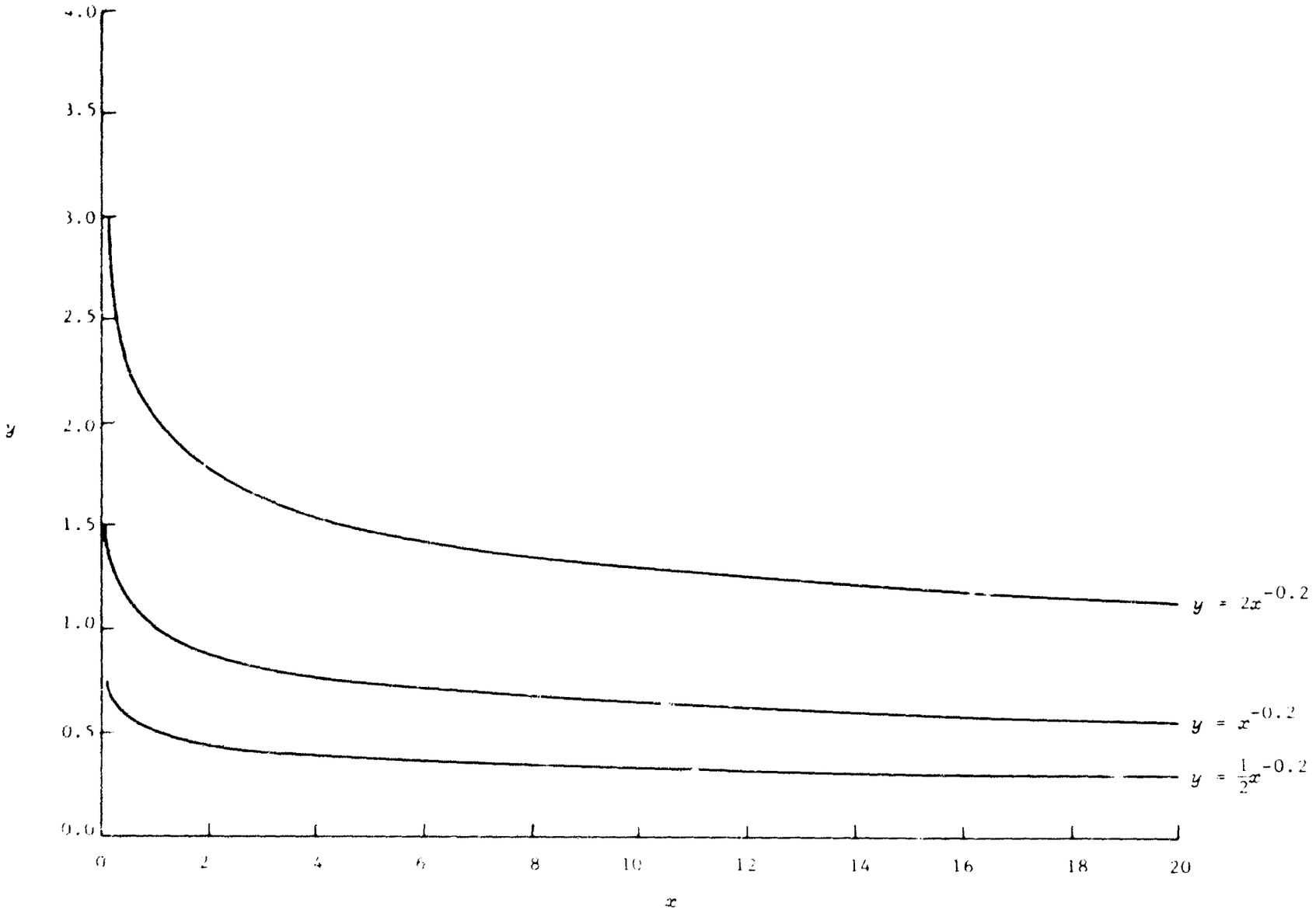


Fig. 21--Power function $y = x^n$ with negative exponent showing effect of x

In most cost analysis applications only that part of the curve where x is > 1 is of interest. Since we are generally concerned with depicting a cost-quantity relationship with y the cost and x the quantity, we are not concerned with the cost of less than one unit. But both because the curve might be useful for other applications and because of its special behavior, we should become familiar with its more general characteristics.*

The power function also is linear when expressed in terms of logarithms. Returning to the general equation

$$y = ax^b,$$

and taking the logarithms of both sides, we have

$$\log y = \log a + b \log x,$$

which is quite clearly a linear expression in logarithms. As most curve-fitting techniques are simpler when handling linear relationships, it is common to make this transformation before fitting the power function. Figure 22 shows the complete family of power functions plotted using logarithmic coordinates.

When two lines (power functions) are parallel on logarithmic coordinates, they differ from each other by a constant ratio or percentage, which is contrary to the case where two parallel lines on arithmetic coordinates differ by a constant number. To demonstrate analytically, assume that we have two curves, one m percent higher than the other. The equations for these curves are

---

*It should also be pointed out that this curve is undefined for a negative b and $x = 0$. In this case y goes to infinity in a positive direction.

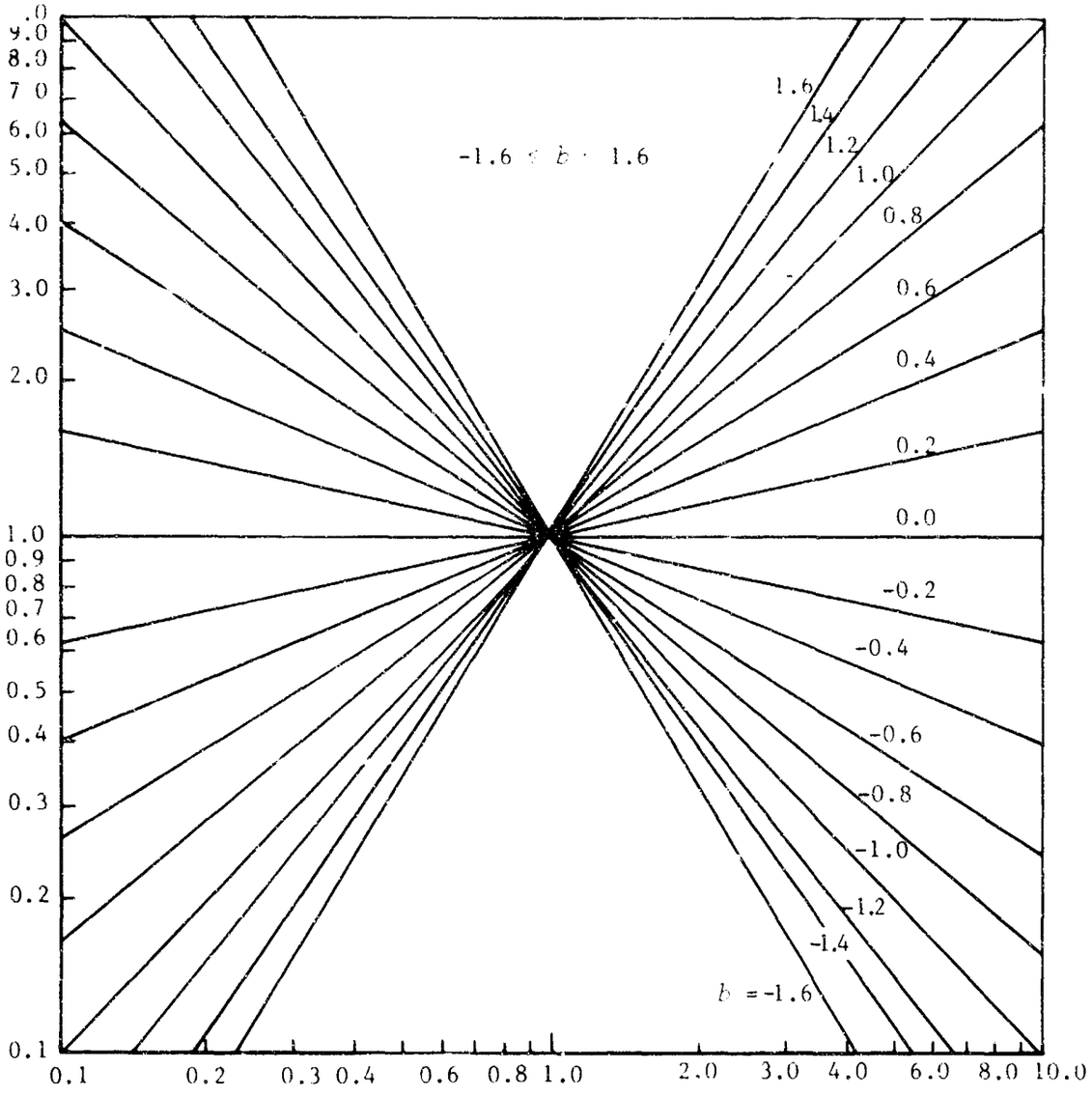


Fig. 22--Power function $y = x^b$ on logarithmic coordinates

$$y_1 = ax^b,$$

$$y_2 = 1\ 5ax^b,$$

and in logarithmic form

$$\log y_1 = \log a + b \log x,$$

$$\log y_2 = \log 1.5 + \log a + b \log x.$$

In the second equation let the constant terms log 1.5 + log a be set equal to log a'. We then have

$$\log y_1 = \log a + b \log x,$$

$$\log y_2 = \log a' + b \log x.$$

The only difference between the two occurs in the constant terms log a and log a'. When these two equations are plotted on logarithmic grids there will be two parallel lines with the spacing between them equal to log a' - log a or log 1.5.

Fitting the power function through two points can be accomplished by transforming both variables into logarithms and proceeding as with a linear case. To review the method, see the discussion of the straight line.*

---

*p. 3ff.

II. FITTING CURVES TO TWO-VARIABLE RELATIONSHIPS

THE STRAIGHT LINE

Although there are many methods of fitting straight lines, three are usually sufficient for fitting curves to two-variable relationships: the method of selected points, the method of averages, and the method of least squares. Of course one can always draw the curves freehand, but even so the equation of the line must be determined by using one of the other methods.

The Method of Selected Points

When it is apparent that data plotted on rectangular coordinates can be described by a straight line, the equation of that line can be found using the method of selected points. With the use of a straight edge, a line is drawn through the points such that the points are uniformly distributed around the line. Two points are then read from the line near the extremities and substituted in the equation

$$y = a + bx.$$

The two equations are then solved simultaneously for a and b.

In the example shown in Fig. 23 the two points selected were P_1 (4, 8.3) and P_2 (26, 24). The two equations were therefore

$$8.3 = a + 4b,$$

$$24.0 = a + 26b.$$

When the first is subtracted from the second, the result is

$$15.7 = 22b,$$

$$b = 0.714.$$

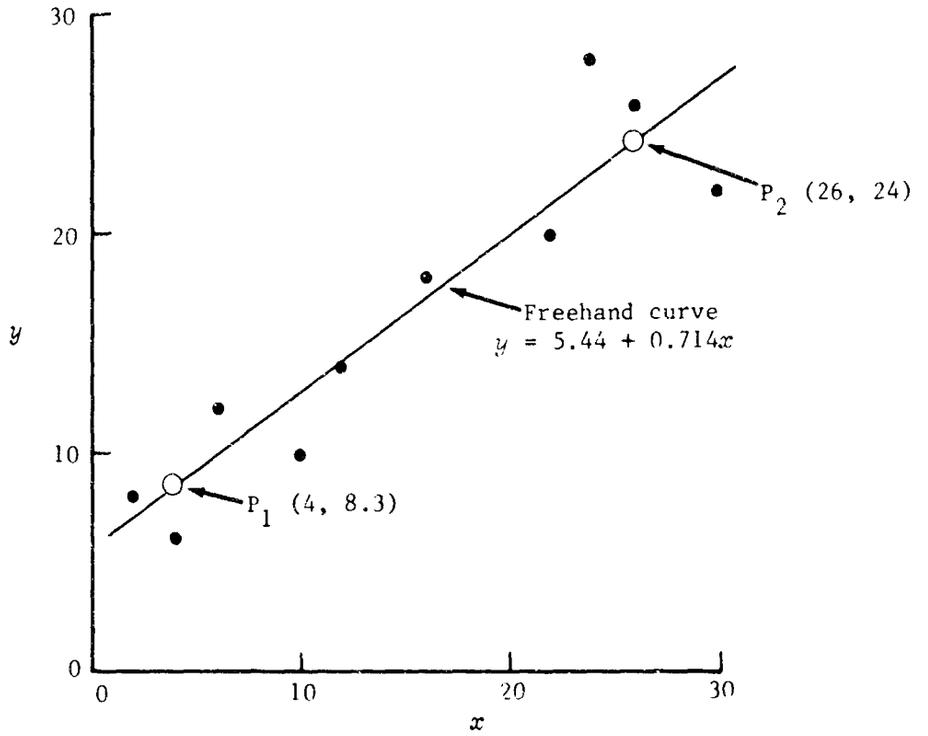


Fig. 23--Straight line fitted usi the method of selected points

When b is substituted in the first equation

$$8.3 = a + (4)(0.714),$$

$$a = 5.44.$$

The equation of the desired line is

$$y = 5.44 + 0.714x.$$

When values of x taken from the original data are substituted in this equation, values of y corresponding to each of the original

values can be calculated. One way of showing how well the line fits the data is to compare the calculated values with the original values using a percent deviation for each data point as in Table 1. When computing the percent deviation, it is usual to base the percentages on the observed values.* For example, the percent deviation for the first data point is

$$(8.00 - 6.87)\frac{100}{8.00} = 14.1.$$

Table 1

USING THE METHOD OF SELECTED POINTS TO FIT A STRAIGHT LINE: DATA AND RESULTS

y	x	y(calc.)	Percent Deviation
8	2	6.87	14.1
6	4	8.30	-38.3
12	6	9.72	19.0
10	10	12.58	-25.8
14	12	14.01	-0.1
18	16	16.86	6.3
20	22	21.15	-5.8
28	24	22.58	19.4
26	26	24.00	7.7
22	30	26.86	-22.1
			15.9 av

After the percent deviations are calculated for each data point, an average percent deviation can be calculated by adding each deviation, disregarding the sign, and dividing the total by the number of data points.

Notice that the placement of the line in the example was quite arbitrary; this is one of the weaknesses of the method of selected

* Percent deviation = $\left(\frac{y \text{ observed} - y \text{ calculated}}{y \text{ observed}}\right) 100.$

points. The same equation could have been arrived at in a slightly different manner. Had the line been extended to the left until it intercepted the y axis, we could have read the value of a directly from the graph and calculated the slope as follows:

$$b = (24 - 8.3)/(26 - 4) = 0.714.$$

Still another method using the two points given and the modified point slope formula would be

$$y - y_1 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right) \cdot (x - x_1),$$

$$y - 8.3 = \left(\frac{24 - 8.3}{26 - 4}\right) \cdot (x - 4),$$

$$y = 5.44 + 0.714x.$$

The Method of Averages

To use the method of averages the data must first be arrayed in ascending order according to one of the variables. Second, the numbers are divided so that two approximately equal groups are formed. If the number of data points is even, there should be an equal number in each; if odd, the extra point will have to be placed in one group or the other. The average value of each of the variables is calculated for each group and substituted into the equation

$$y = a + bx.$$

Two equations result as before; these are solved simultaneously for a and b.

In the example shown in Fig. 24 and Table 2, the data are ordered according to the variable x. As there are an even number of data points (10), 5 are to be assigned to each group. When average values for both x and y are calculated for each group, they define the two points (6.8, 10) and 23.6, 22.8) which are then plotted and a straight line connecting them drawn. The equation of the line is obtained by substituting the coordinates of the average points into the equation $y = a + bx$ as before and solving the two equations simultaneously for a and b.

$$10 = a + 6.8b,$$

$$22.8 = a + 23.6b.$$

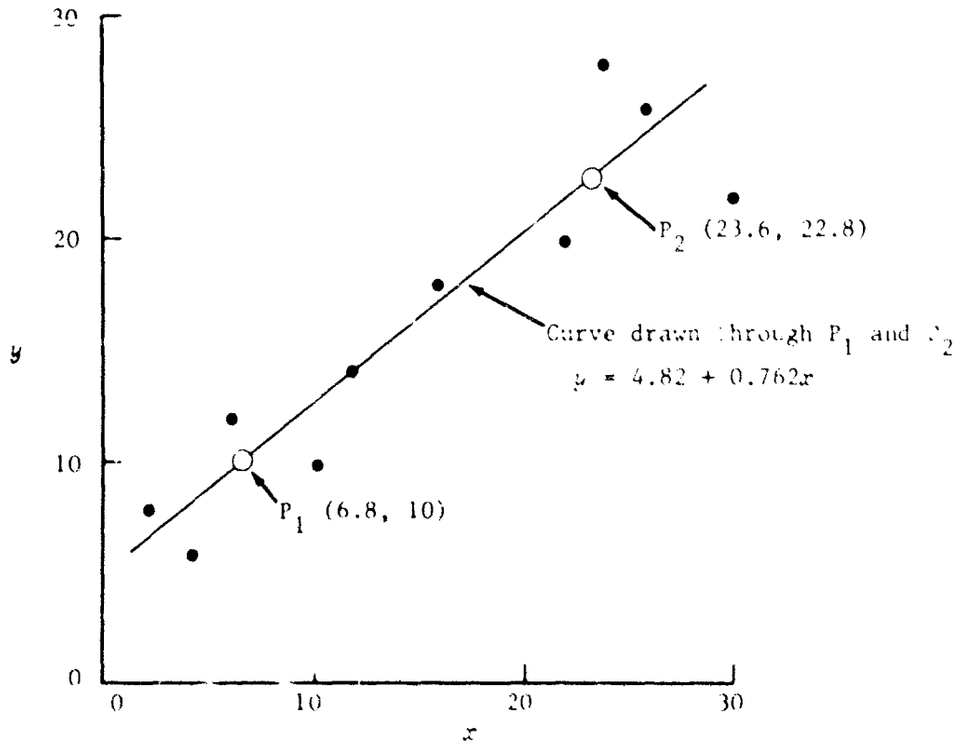


Fig. 24--Straight line fitted using the method of averages

When the first is subtracted from the second, the result is

$$12.8 = 16.8b,$$

$$b = 0.762;$$

and substituting b in the first equation,

$$10 = a + (6.8)(0.762),$$

$$a = 4.82.$$

Therefore,

$$y = 4.82 + 0.762x.$$

Percent and average percent deviations as calculated are shown in Table 2.

Table 2

USING THE METHOD OF AVERAGES TO FIT A STRAIGHT LINE: DATA AND RESULTS

y	x	y (calc.)	Percent Deviation
8	2	6.3	20.7
6	4	7.87	-31.2
12	6	9.39	21.7
10	10	12.44	-24.4
14	12	13.96	0.3
10 av	6.8 av	--	--
18	16	17.01	5.5
20	22	21.58	-7.9
28	24	23.11	17.5
26	26	24.63	5.3
22	30	27.68	-25.8
22.8 av	23.6 av	--	16.0 av

A slightly different version of the method of averages is sometimes used. Instead of calculating average values as was done before, the data are used to establish ten separate linear equations as follows:

$8 = a + 2b$	$18 = a + 16b$
$6 = a + 4b$	$20 = a + 22b$
$12 = a + 6b$	$28 = a + 24b$
$10 = a + 10b$	$26 = a + 26b$
$14 = a + 12b$	$22 = a + 30b$
$50 = 5a + 34b$	$114 = 5a + 118b$

These groupings are preserved and the equations appearing in each group are added to obtain the required two equations which are solved simultaneously for a and b:

$$50 = 5a + 34b,$$

$$114 = 5a + 118b.$$

Subtracting the first from the second,

$$64 = 84b,$$

$$b = 0.762;$$

and substituting b in the first equation

$$50 = 5a + (34)(0.762),$$

$$a = 4.82.$$

The same solution results from either averaging or adding. Although the method of averages is simple to use and does give a reproducible

solution does *not* ensure that a *best* fitting straight line will be chosen.

The Method of Least Squares

The method of least squares is probably the most widely used method of obtaining empirical equations. The term "least squares" reflects the criterion used to determine the desired equation. The line is chosen such that the sum of the squared deviations of the points from the line is minimized.* The way this criterion is used in the formula for calculating a least-squares line is worked out in full in Appendix A. Briefly it is as follows:

We are seeking an equation of a straight line

$$y = a + bx,$$

such that the sum of the squared distances of the data from that line will be minimal.

The sum of the squared deviations (D) is expressed mathematically as

$$D = \sum (y - a - bx)^2,$$

To obtain values of a and b so that this expression can be minimized, it is necessary to take partial derivatives with respect to both a and b and to equate them to 0.** These partial derivatives are

$$\frac{\partial D}{\partial a} = -2\sum y + 2Na + 2b\sum x = 0, \text{ and}$$

*Where the equation is of the form $y = a + bx$, the distances are measured parallel to the y axis. Conversely, where the equation is of the form $x = a + by$, the distances are measured parallel to the x axis.

**This is a method of calculus included here only as a matter of interest. Its comprehension is not essential to anything that follows in this text.

$$\frac{\partial \Sigma d^2}{\partial b} = -2 \sum xy + 2a \sum x + 2b \sum x^2 = 0;$$

and with the obvious simplifications made, the resultant two equations

$$\sum y = Na + b \sum x,$$

$$\sum xy = a \sum x + b \sum x^2.$$

are called the normal equations for fitting a least-squares straight line. Each of the values other than a and b can readily be determined from the data, leaving two equations with two unknowns that can be solved simultaneously.

In applying this method it is convenient to array the data as in the first two columns of Table 3. (The ordering of the values is not essential, although it tends to make checking the calculations easier.) The next step is to square each entry in the column headed x and enter

Table 3

USING THE METHOD OF LEAST SQUARES TO FIT A STRAIGHT LINE: WORKSHEET

y	x	x^2	xy	y(calc.)	d	Percent Deviation	d^2
8	2	4	16	7.07	0.93	11.7	0.865
6	4	16	24	8.49	-2.49	-41.5	6.200
12	6	36	72	9.90	2.10	17.5	4.410
10	10	100	100	12.73	-2.73	-27.3	7.453
14	12	144	168	14.14	-0.14	-1.0	0.020
18	16	256	288	16.97	1.03	5.7	1.061
20	22	484	440	21.21	-1.21	-6.1	1.464
28	24	576	672	22.63	5.37	19.2	28.837
26	26	676	676	24.04	1.96	7.5	3.842
22	30	900	660	26.87	-4.87	-22.1	23.717
164	152	3,192	3,116	--	--	16.0 av.	77.869

the product in the column headed x^2. The entries in the column headed xy are the products of the x and y values for each point. After these calculations have been made, each column is totaled as shown. N is the number of data points; the remaining values required by the normal equations are the totals of the appropriate columns. When these are substituted in the normal equations, we have

$$164 = 10a + 152b,$$

$$3116 = 152a + 3192b,$$

and when these are solved simultaneously,

$$a = 5.66$$

$$b = 0.707.$$

The equation is therefore

$$y = 5.66 + 0.707x.$$

Figure 25 shows the data and the straight line described by this equation. The average percent deviation is calculated as before. Another measure of goodness of fit that is often used, particularly in conjunction with least-squares, is the standard error of the estimate of y, S.* This measure is obtained by squaring each of the deviations, adding the results, dividing the total by N (the number of data points), and taking the square root of the result, e.g.,

$$S_y = \sqrt{\frac{\Sigma d^2}{N-2}}.$$

*The standard error of the estimate of y allows a heavier penalty for extreme data points than does the average percent deviation method, yet it is more difficult to interpret.

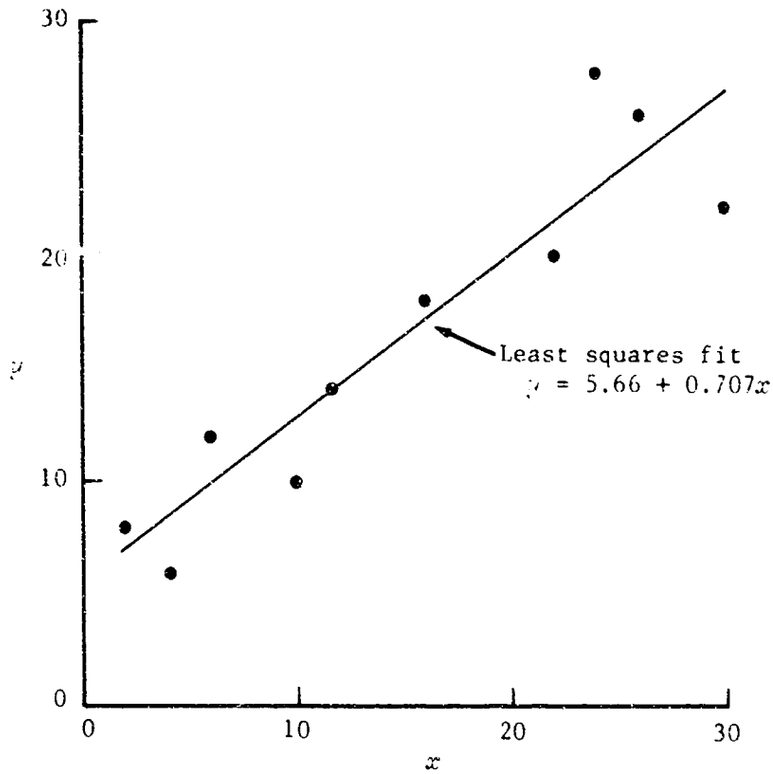


Fig. 25--Straight line fitted using the method of least squares

In our example

$$S = \pm\sqrt{\frac{77.869}{8}},$$

$$S = \pm 3.120.$$

Since the standard error of the estimate of y is expressed in the same units as the variable y, it is often better, when making comparisons, to convert to a non-dimensional number. This is usually done by dividing S by the arithmetic mea.. of y, $\bar{y}$:

$$V = (S/\bar{y})100,$$

or

$$V = \frac{3.12}{16.4} \times 100;$$

$V = 19.02\%$,

where V is called the coefficient of variation.

Summary

Of the three methods for fitting straight lines the method of selected points is the easiest to use, but when the required line is not obvious from the data, the choice is strictly a matter of judgment and the results are not easily reproducible. The method of averages provides a relatively simple and yet definitive way of choosing an equation (although not necessarily the best fitting line). The only place where uncertainty enters the picture is in the placement of the odd data point. The method of least squares requires more in the way of calculation but provides a completely unambiguous way of selecting and fitting the line; for that and other reasons it is the most widely used method.

An average percent deviation can be used to show how well the line fits the data. It is easy to calculate and to interpret. The standard error of the estimate of y, while requiring more calculations to be made, is more widely used because of its implications in drawing statistical inferences. The coefficient of variation, a non-dimensional term based on the standard error, is useful especially for making comparisons.

Although the results of applying the three different methods to the same data, as displayed in Fig. 26, are strikingly similar, the user cannot expect that this would always be the case.

THE PARABOLA

The same three methods used to fit the straight line may be used to fit the parabola. To fit the parabola, both the method of selected

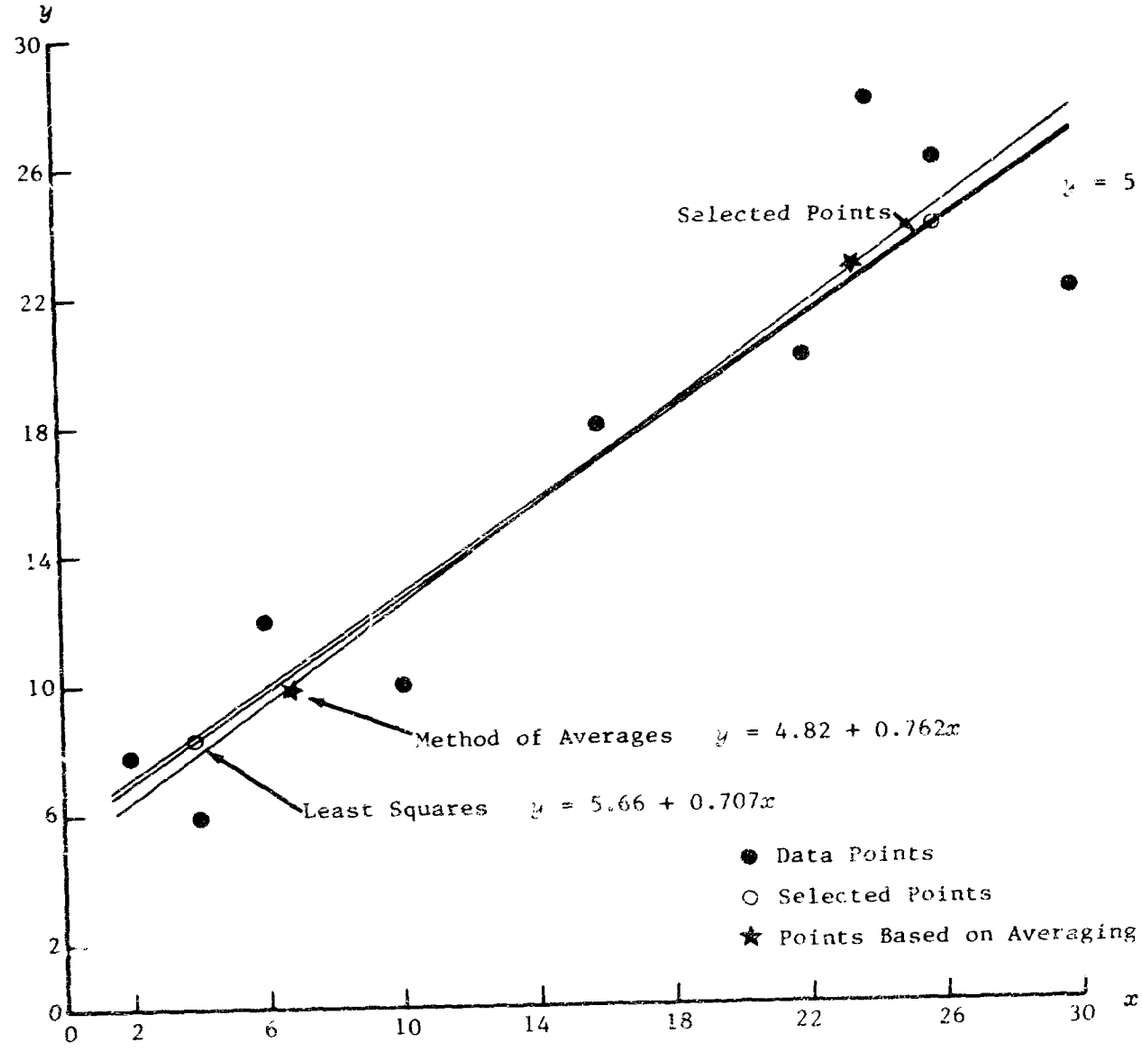


Fig. 26--Results of fitting a straight line to the same data using three alternate methods

points and the method of averages rely even more on the judgment of the cost analyst than was the case with respect to the straight line. For this and other reasons, the method of least squares is preferred. Measuring goodness of fit is the same as it was for the straight line and will not be discussed again. There are two forms of the parabola which may be used by the cost analyst. The first is

$$y = a + bx + cx^2,$$

which is illustrated in Fig. 27. The second is

$$x = a + by + cy^2,$$

or, in terms of y,

$$y = \frac{-b \pm \sqrt{b^2 - 4c(a - x)}}{2c},$$

which is illustrated in Fig. 28.

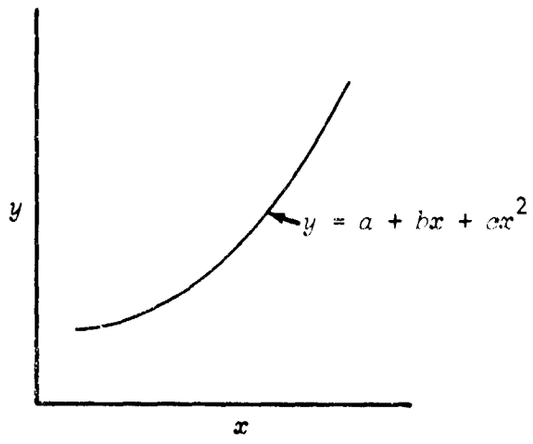


Fig. 27--Parabola form 1

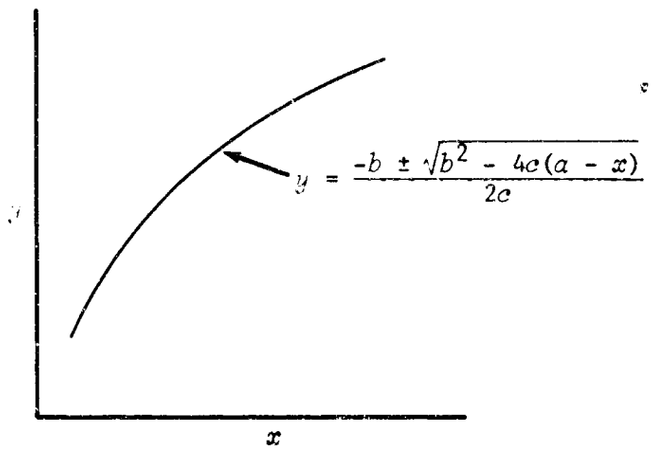


Fig. 28--Parabola form 2

In the first case, the line of symmetry of the parabola is parallel to the y axis and in the second it is parallel to the x axis. Each form will be examined separately. Form 2 requires y to be the squared term. In order to accomplish this, the equation of the first form is written interchanging y and x as follows:

$$x = a + by + cy^2;$$

and solving for y using the quadratic formula

$$cy^2 + by + a - x = 0,$$

$$y = \frac{-b \pm \sqrt{b^2 - 4c(a - x)}}{2c}.$$

Parabola Form 1

The Method of Selected Points. When it is clear that the relationship described by the data can be represented by a segment of a parabola, the method of selected points may be used. First, draw a freehand curve roughly the shape of a parabola through the data, and

read three points from the curve, two at the extremities and one somewhere in the middle. If there is a relatively sharp maximum or minimum point (vertex), it would be best to read the third point from the area of the vertex. Then substitute three points in the equation

$$y = a + bx + cx^2,$$

and solve the resulting three equations simultaneously for a, b, and c. Notice that x appears twice in each equation, once as it is read from the curve and once squared.

In the example shown in Fig. 29, numbers for which are given in Table 4, we draw the freehand curve indicated by the solid line and read from

Table 4

USING THE METHOD OF SELECTED POINTS TO FIT THE PARABOLA FORM 1: DATA AND RESULTS

y	x	y (calc)	Percent Deviation
3	1	2.00	-33.3
2	3	2.99	50.0
4	4	3.56	-11.0
5	8	6.37	27.4
7	10	8.08	15.4
10	11	9.01	-9.9
12	14	12.13	1.1
15	17	15.70	4.7
18	18	17.00	-5.6
17	19	18.35	7.9
--	--	--	16.6 av

the curve the points (18, 17), (12, 10) and (1, 2). The required three equations are

$$17 = a + 18b + 324c,$$

$$10 = a + 12b + 144c,$$

$$2 = a + b + c.$$

When we subtract the second from the first and the third from the second, we get

$$7 = 6b + 180c,$$

$$8 = 11b + 143c.$$

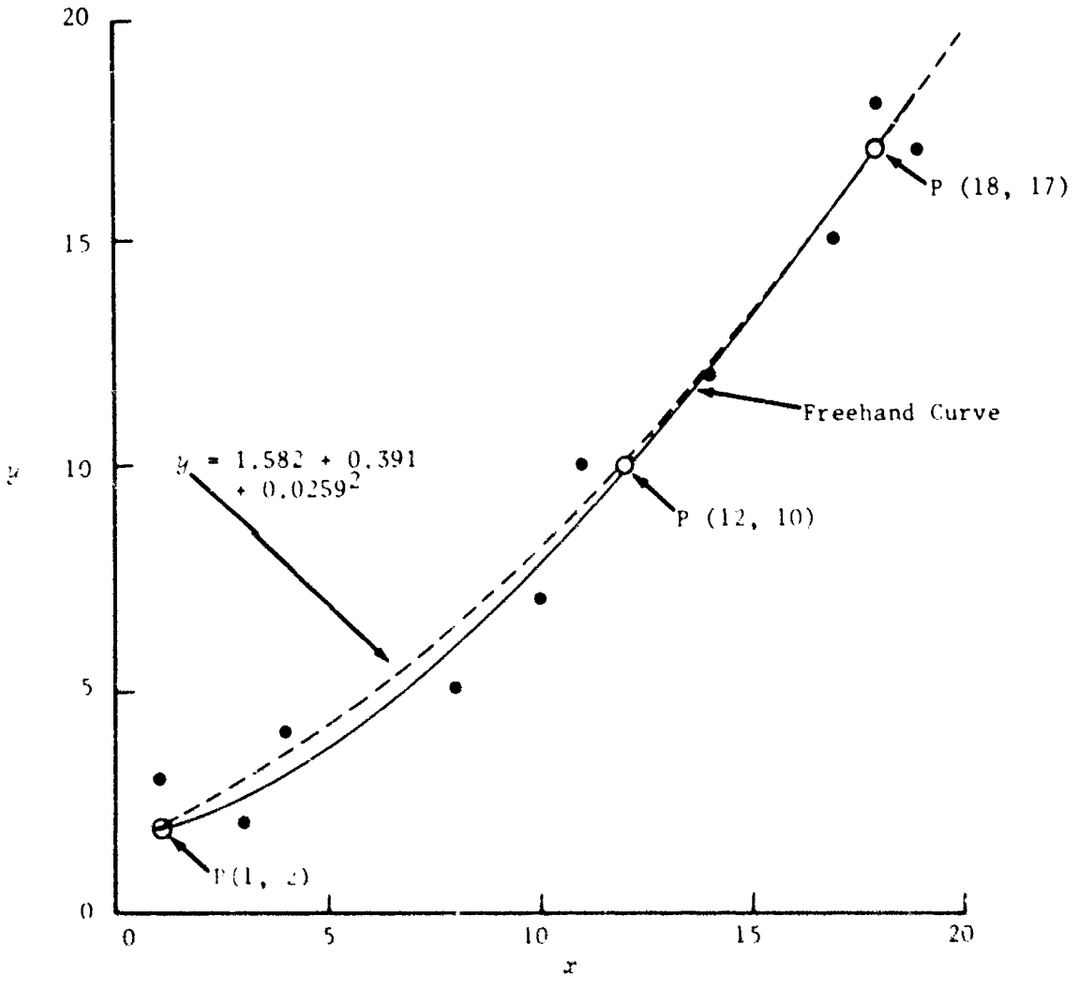


Fig. 29--Parabola form 1 fitted using the method of selected points

To eliminate b, we multiply the first equation by 11 and the second by 6, then subtract the second from the first:

$$29 = 1122c,$$

$$c = 0.0259.$$

Substituting c in the first equation,

$$7 = 6b + (180)(0.0259),$$

$$b = 0.391.$$

Next, we substitute both b and c in the third of the original equations, and calculate a as follows:*

$$2 = a + b + c,$$

$$a = 2 - 0.391 - 0.0259,$$

$$a = 1.582.$$

The required equation is therefore

$$y = 1.582 + 0.391x + 0.0259x^2.$$

The graph of this equation is shown by the dashed line in Fig. 29.

As with the straight line, the chances are small that two analysts independently using this method would arrive at the same result. The only argument in favor of it is that it is relatively simple to use.

The Method of Averages. Three points similar to those obtained in the method of selected points by arbitrarily choosing them from a

*The discrepancies in the arithmetic result from the fact that more decimal places than those shown were used in making the actual calculations.

freehand curve may be obtained by averaging. In this method, we array the data in ascending order of one of the variables, form three groups of approximately equal size, and calculate the average values of both x and y for each group. We substitute the average points in the equation

$$y = a + bx + cx^2,$$

and solve the resulting three equations simultaneously for a, b, and c.

Table 5 illustrates the procedure for the example case. The 10 numbers are arrayed in ascending order according to the value of x. Three groups are formed by assigning 3 points to the first and last groups and 4 points to the middle one.

Table 5

USING THE METHOD OF AVERAGES TO FIT THE PARABOLA FORM 1: WORKSHEET

y	x	y (calc)	Percent Deviation
3	1	2.34	22.0
2	3	3.15	-57.5
4	4	3.64	9.0
3 av	2.67 av	--	--
5	8	6.20	-24.0
7	10	7.83	-11.9
10	11	8.73	-12.7
12	14	11.78	1.8
8.5 av	10.75 av	--	--
15	17	15.36	-2.4
18	18	16.67	7.4
17	19	18.04	-6.1
16.67 av	18 av	--	15.5 av

(The assignment of the odd point is arbitrary.) By averaging, we obtain

three points, (18, 16.67), (10.75, 8.5), and (2.67, 3), and plot them as shown in Fig. 30.

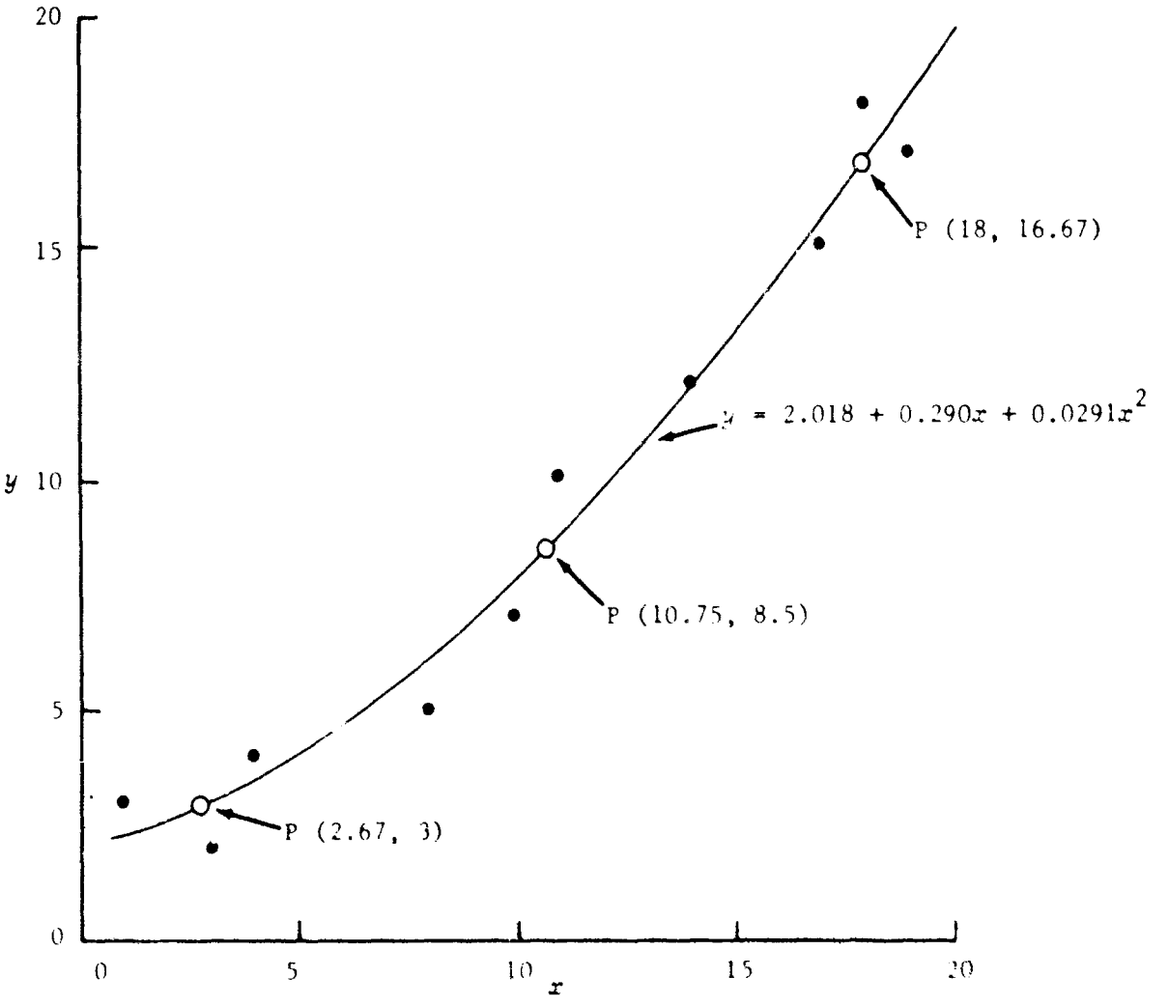


Fig. 30--Parabola form 1 fitted using the method of averages

We substitute these same points in the equation

$$y = a + bx + cx^2,$$

which results in the three following equations:

$$3.00 = a + 2.67b + 7.129c,$$

$$8.50 = a + 10.75b + 115.563c,$$

$$16.67 = a + 18.00b + 324.000c.$$

When these equations are solved simultaneously, the result is

$$a = 2.018,$$

$$b = 0.290,$$

$$c = 0.0291.$$

The equation of the required parabola is therefore

$$y = 2.018 + 0.290x + 0.0291x^2.$$

This method is less arbitrary than the method of selected points, but some ambiguity does exist because of having to assign any odd data point to one of the three groups. Further, averaging may prevent us from choosing a point near the vertex of the curve.

The Method of Least Squares. To fit a parabola using the method of least squares, we must solve three normal equations simultaneously for the values of the coefficients a, b, and c in the equation

$$y = a + bx + cx^2.$$

The normal equations are

$$\sum y = Na + b \sum x + c \sum x^2,$$

$$\sum xy = a \sum x + b \sum x^2 + c \sum x^3,$$

$$\sum x^2y = a \sum x^2 + b \sum x^3 + c \sum x^4.$$

The least-squares criterion and details of the derivation of these equations, which are described fully in the Appendix, are similar to those for the straight line. As in previous discussions, all of the information required to solve the equations can be obtained from the data.

In a worksheet similar to the one shown in Table 6, we array the data as in the first two columns, making entries in the other columns after performing the calculations indicated by the headings. N is the

Table 6

USING THE METHOD OF LEAST SQUARES TO FIT
THE PARABOLA FORM 1: WORKSHEET

y	x	x^2	x^3	x^4	xy	x^2y	y (calc)	Percent Deviation
3	1	1	1	1	3	3	2.28	24.0
2	3	9	27	81	6	18	3.04	-52.0
4	4	16	64	256	16	64	3.51	12.0
5	8	64	512	4,096	40	320	6.03	-20.6
7	10	100	1,000	10,000	70	700	7.66	-9.4
10	11	121	1,331	14,641	110	1,210	8.58	14.2
12	14	196	2,744	38,416	168	2,352	11.69	2.6
15	17	289	4,913	83,521	255	4,335	15.37	-2.5
18	18	324	5,832	104,976	324	5,832	16.72	7.1
17	19	361	6,859	130,321	323	6,137	18.13	-6.6
93	105	1,481	23,283	386,309	1,315	20,971	---	15.1 av

number of data points and the column totals provide the other necessary inputs to the normal equations. The equations to be solved for the example shown in Fig. 31 and Table 6 are:

$$93 = 10a + 105b + 1481c,$$

$$1315 = 105a + 1481b + 23283c,$$

$$20971 = 1481a + 23283b + 386309c.$$

Therefore

$$a = 1.993,$$

$$b = 0.254,$$

$$c = 0.0314.$$

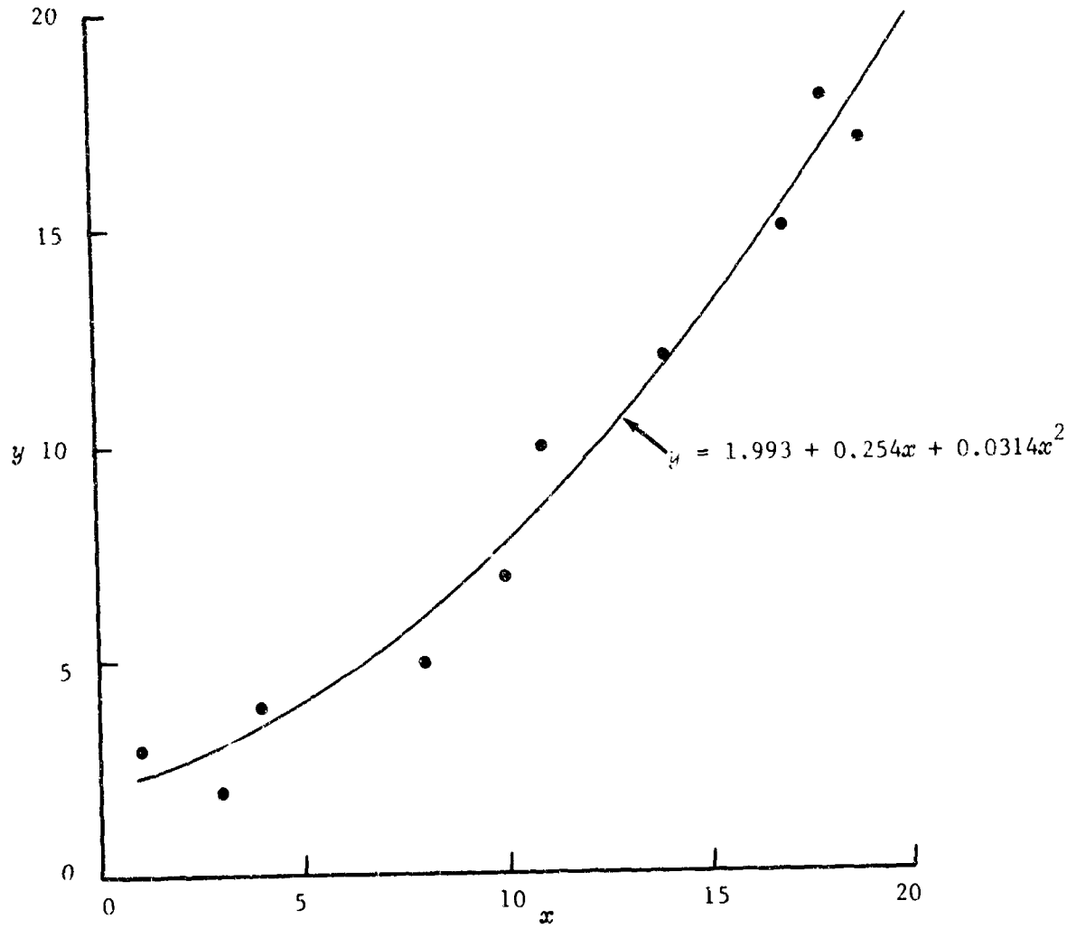


Fig. 31--Parabola form 1 fitted using the method of least squares

Thus the desired equation is

$$y = 1.993 + 0.254x + 0.0314x^2.$$

The method of selected points, the method of averages, and the method of least squares can each be used to fit a parabola. However, because the method of least squares results in an unambiguous solution, it is usually preferred.

Figure 32 shows a comparison of the results obtained by using each of these methods to fit a parabola to the example data.

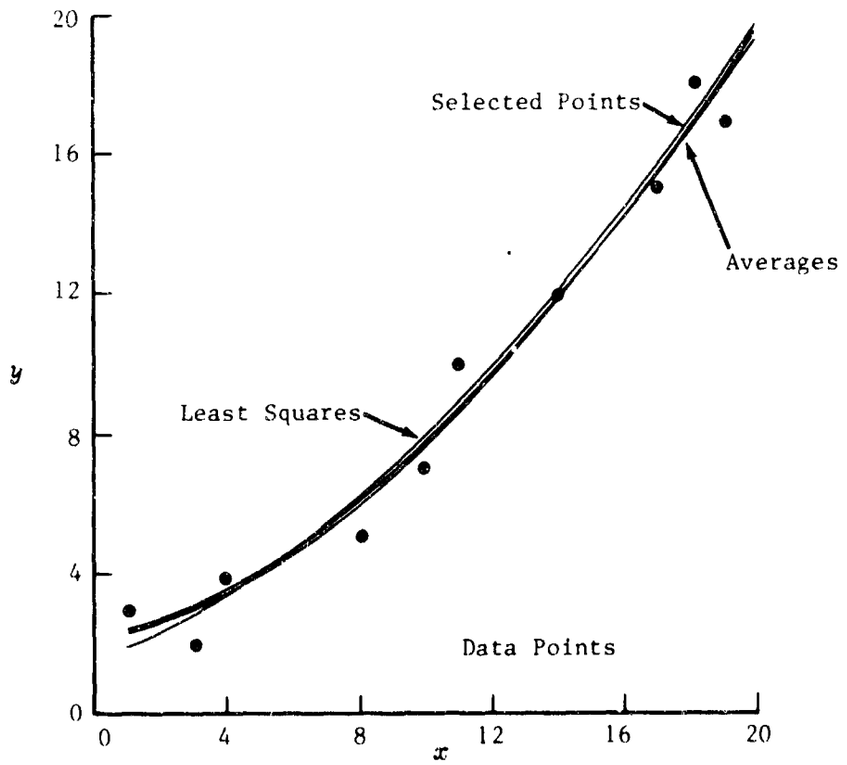


Fig. 32--Fitting a parabola form 1 using three alternate methods

Parabola Form 2

The equation for this class of parabolas is

$$x = a + by + cy^2.$$

The only difference between this and the equation for the parabola form 1 is that x and y have been interchanged. In fitting this form we invert the relationship between x and y in the data and proceed as before until a, b, and c have been calculated. At that point the equation in y as above must be solved for y using the quadratic formula. The desired result will typically be one of the two possible solutions; the appropriate one can best be determined by experimentation. While all three curve fitting methods can be used here also, we shall only illustrate the method of least squares.

The Method of Least Squares. The normal equations necessary to fit this kind of parabola are the same as for form 1 but with x and y interchanged as

$$\sum x = Na + b \sum y + c \sum y^2,$$

$$\sum xy = a \sum y + b \sum y^2 + c \sum y^3,$$

$$\sum xy^2 = a \sum y^2 + b \sum y^3 + c \sum y^4.$$

When the appropriate values are calculated as shown in Table 7 and substituted in the above equations, we have

$$92 = 10a + 110b + 1530c,$$

$$1370 = 110a + 1530b + 23690c,$$

$$22076 = 1530a + 23690b + 387858c.$$

Table 7

USING THE METHOD OF LEAST SQUARES TO FIT
THE PARABOLA FORM 2: WORKSHEET

y	x	y^2	y^3	y^4	xy	xy^2	y (calc)	Percent Deviation
2	1	4	8	16	2	4	0.76	62.0
4	2	16	64	256	8	32	3.99	0.3
6	3	36	216	1,296	18	108	5.80	3.3
8	5	64	512	4,096	40	320	8.40	-5.0
10	9	100	1,000	10,000	90	900	12.13	-21.3
12	6	144	1,728	20,736	72	864	9.46	21.2
15	12	225	3,375	50,625	180	2,700	14.33	4.5
16	16	256	4,096	65,536	256	4,096	16.84	-5.3
18	18	324	5,832	104,976	324	5,832	17.97	0.2
19	20	361	6,859	130,321	380	7,220	19.04	-0.2
110	92	1,530	23,690	387,858	1,370	22,076	---	12.3 av

The solution is

$$a = 0.913,$$

$$b = 0.0783,$$

$$c = 0.0485;$$

and the equation sought is

$$x = 0.913 + 0.0783y + 0.0485y^2.$$

Since our objective is to use x to estimate y, we must solve this equation for y. We can do this by writing it as a quadratic equation in y:

$$0.0485y^2 + 0.0783y + (0.913 - x) = 0,$$

and using the quadratic formula we obtain

$$y = \frac{-0.0783 \pm \sqrt{0.0783^2 - 4(0.0485)(0.913 - x)}}{2(0.0485)},$$

which simplifies to

$$y = \frac{-0.0783 \pm \sqrt{-0.171 + 0.194x}}{0.0971}.$$

We can see on inspection that the solution in which the square-root term is negative is not useful; the correct equation is

$$y = \frac{-0.0783 + \sqrt{-0.171 + 0.194x}}{0.0971}.$$

This equation has been graphed in Fig. 33. The reason for selecting the form 2 parabola is that as larger and larger values of x are used, the value of y continues to increase--at a decreasing rate, however. Such would not have been true had a form 1 parabola been used. This problem is discussed in the section of this Memorandum on analytic geometry.*

THE EXPONENTIAL

In its simplest form, the equation of the exponential is

$$y = e^x,$$

or

$$y = 10^x,$$

depending on whether base e or 10 is preferred. (Recall the earlier discussion of the advantages of each.)

*See pp. 19-22.

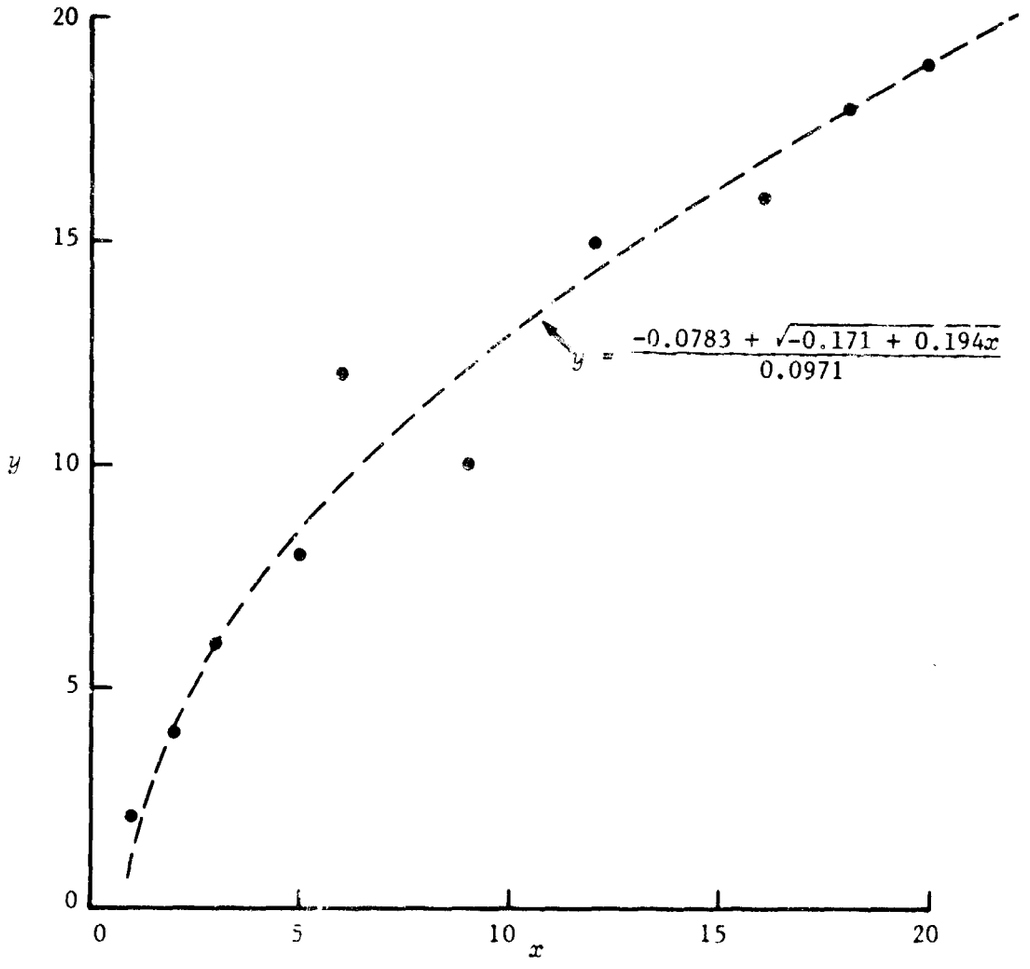


Fig. 33--Parabola form 2 fitted using the method of least squares

For our purposes, a more useful form of the exponential is

$$y = e^{a+bx},$$

or

$$y = 10^{a+bx}.$$

The latter equation allows the y intercept to take on values other than 1, depending on the value of a, and to accelerate at a rate

greater or less than x, depending on the value of b. For illustrative purposes, we will work exclusively with

$$y = 10^{a+bx},$$

although those who prefer to use base e may do so, since the procedures are the same.

Unfortunately there is no direct least-squares solution for fitting a curve of this type. There are iterative methods that can be used to approximate a least-squares solution, but they require a large computer to be of practical use.*

The usual method is to transform the exponential into a linear equation by taking the logarithms of each side as follows:

$$\log y = a + bx.$$

We then substitute the logarithms of the y values for the actual values and employ the least-squares normal equations for fitting a straight line. It should be noted, however, that this method does not yield the same least-squares solution for a and b as the exponential form does. The criterion of least squares is applied to the logarithms, not to the actual values of y, which results in minimization of the relative rather than the absolute deviations.** The fitted line is also higher than would be the case had the least-squares criterion been

---

*C. A. Graver and H. E. Boren, *Multivariate Logarithmic and Exponential Regression Models*, The RAND Corporation, RM-4879-PR, July 1967; in this Memorandum, the term "exponential" applies to the power form used in this text.

**Ibid. This approach is fine when one wants to minimize relative rather than absolute differences. One could argue that such is the case for most cost-analysis problems.

applied to the actual values. A similar phenomenon occurs when the method of averages is used.

To illustrate the least-squares method we apply it to the data in Table 8. Notice that the first step is to obtain the logarithms of the y values. From that point on, the calculations required are as indicated in the column headings. The normal equations are the same as for the linear case with $\log y$ substituted for y:

$$\sum \log y = Na + b \sum x$$

$$\sum x \log y = a \sum + b \sum x^2.$$

Table 8

USING THE METHOD OF LEAST SQUARES TO FIT THE EXPONENTIAL: WORKSHEET (semi-log form)

y	$\log y$	x	x^2	$x \log y$	$\log y$ (calc)	y (calc)	Percent Deviation
30	1.4771	1	1	1.477	1.387	24.4	18.7
19	1.2788	2	4	2.558	1.280	19.0	0.0
15	1.1761	3	9	3.528	1.172	14.9	0.7
10	1.0000	4	16	4.000	1.065	11.6	-16.1
9	0.9542	5	25	4.771	0.958	9.1	-1.1
6	0.7782	6	36	4.669	0.851	7.1	-18.3
5	0.6990	7	49	4.893	0.744	5.5	-10.0
4	0.6021	8	64	4.816	0.636	4.3	-7.5
4	0.6021	9	81	5.419	0.529	3.4	15.0
3	0.4771	10	100	4.771	0.422	2.6	13.3
105	9.0447	55	385	40.902	---	---	10.1 av

Substituting the appropriate values from Table 8 yields

$$9.0446 = 10a + 55b,$$

$$40.902 = 55a + 385b,$$

which when solved result in

$$a = 1.494,$$

$$b = -0.1072.$$

The desired equation is therefore either

$$\log y = 1.494 - 0.1072x,$$

or

$$y = 10^{1.494-0.1072x}.$$

The graph of this solution is shown in Fig. 34. Because only the left-hand member of the equation is expressed in logarithms, this solution is often called the semi-log form.

When the log transformation of y is not entirely sufficient to straighten out the data, adding or subtracting a constant from the value of y may help. The equation that results when the constant is used is

$$y - \alpha = 10^{a+bx},$$

or

$$y = 10^{a+bx} + \alpha,$$

and in semi-log form:

$$\log (y - \alpha) = a + bx.$$

The value of the constant can be found by trial and error, but is more conveniently estimated using the following procedure. The data are plotted as in Fig. 35 (a) and a freehand curve is drawn. Three points

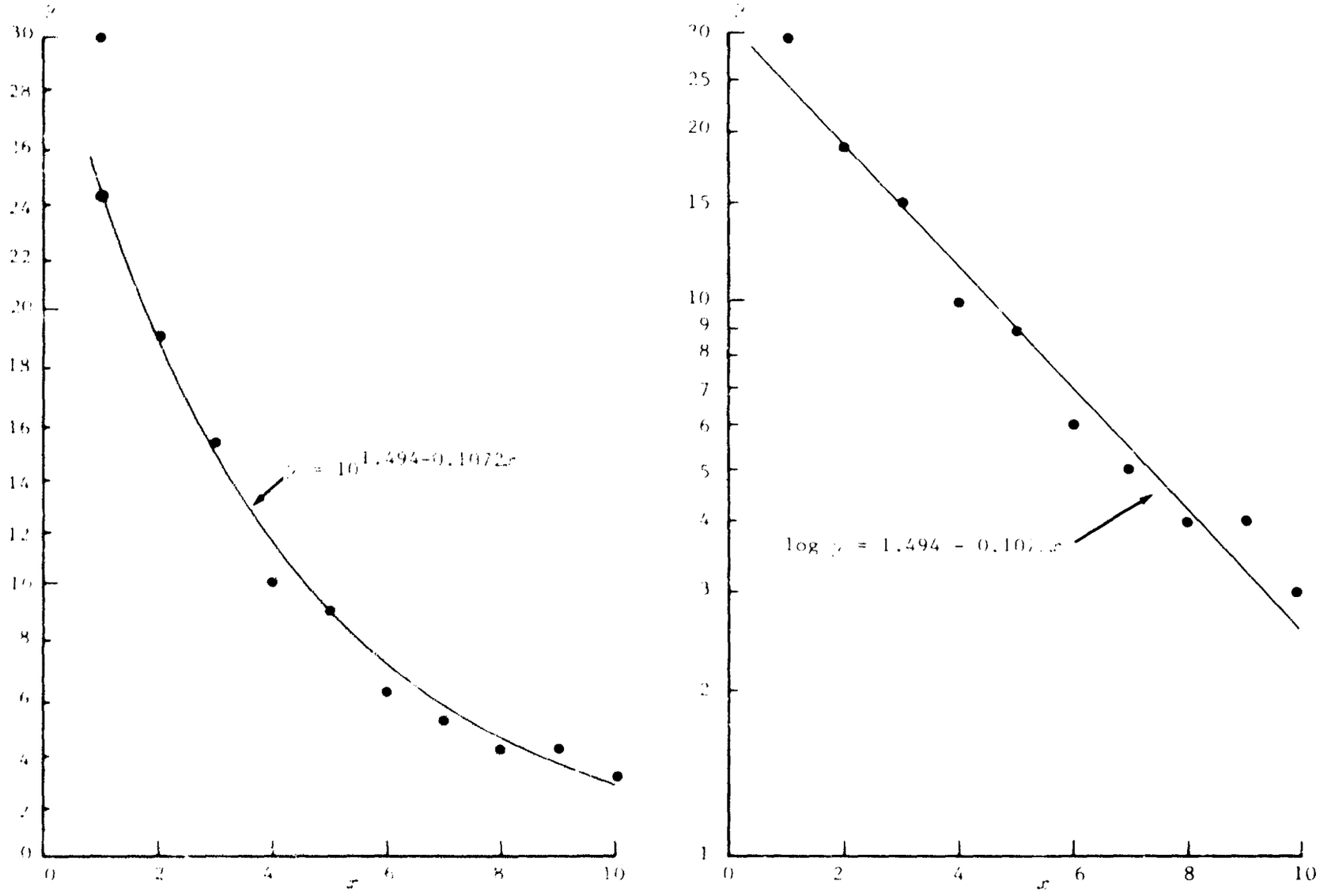


Fig. 34--The exponential fitted using the method of least squares, and the equivalent semi-log form.

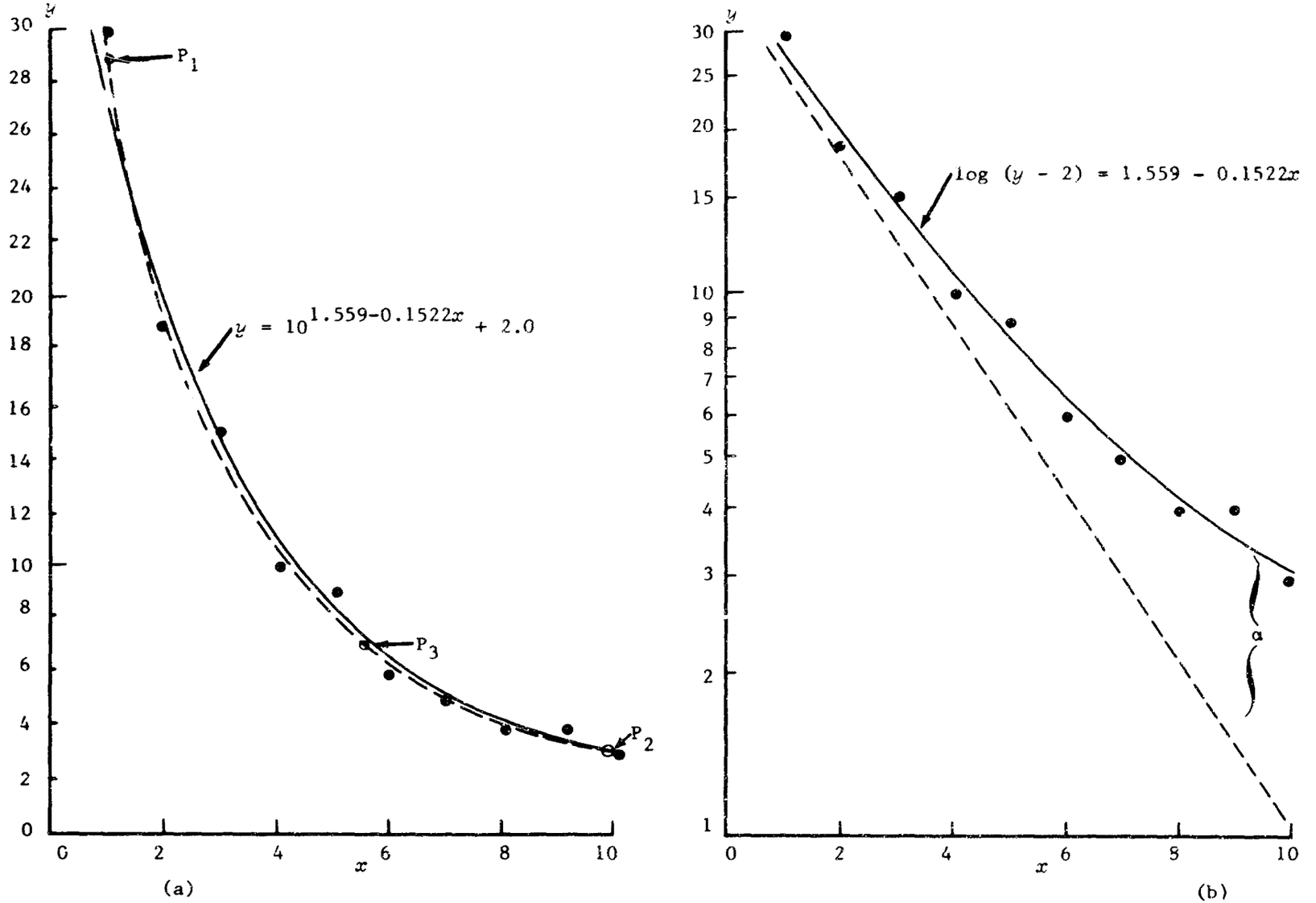


Fig. 35--Determining the constant α using the method of least squares

are selected such that two lie at the extremities of the curve and the third lies halfway between. If the first two points have coordinates x_1, y_1, and x_2, y_2, then x_3 will be equal to $(x_1 + x_2) \div 2$. The y coordinates for each of these points are read from the curve and substituted in the equation

$$\alpha = \frac{y_1 y_2 - y_3^2}{y_1 + y_2 - 2y_3},$$

and α is estimated. See Appendix B for the derivation of this formula.

To illustrate, the three points read from the curve in Fig. 35a are

$$P_1 = (x_1, y_1) = (1, 29),$$

$$P_2 = (x_2, y_2) = (10, 3),$$

$$P_3 = (x_3, y_3) = (5.5, 7.2),$$

and

$$\alpha = \frac{(29)(3) - (7.2)^2}{29 + 3 - (2)(7.2)},$$

$$\alpha = 2.0.$$

The value of α is subtracted from each value of y in the data and the logs of $(y - \alpha)$ are determined. The two steps are shown in Table 9. From that point on, the steps are the same as used in the semi-log or exponential case. When the appropriate values are calculated and the normal equations solved, the results are

$$a = 1.559,$$

$$b = -0.1522,$$

$$\alpha = 2.00,$$

Table 9

USING THE METHOD OF LEAST SQUARES TO FIT THE EXPONENTIAL WITH THE CONSTANT α: WORKSHEET

y	$(y-\alpha)$	$\log(y-\alpha)$	x	x^2	$x \log(y-\alpha)$	$\log(y-\alpha)_c$	$(y-\alpha)_c$	y_c	Percent Deviation
30	28	1.4472	1	1	1.447	1.407	25.5	27.5	8.3
19	17	1.2304	2	4	2.461	1.255	18.0	20.0	5.3
15	13	1.1139	3	9	3.342	1.103	12.7	14.7	2.0
10	8	0.9031	4	16	3.612	0.950	8.9	10.9	-9.0
9	7	0.8451	5	25	4.225	0.798	6.3	8.3	7.8
6	4	0.6021	6	36	3.612	0.646	4.4	6.4	-6.7
5	3	0.4771	7	49	3.340	0.494	3.1	5.1	-2.0
4	2	0.3010	8	64	2.408	0.342	2.2	4.2	-5.0
4	2	0.3010	9	81	2.709	0.189	1.5	3.5	12.5
3	1	0.0000	10	00	0.000	0.037	1.1	3.1	-3.3
105	85	7.2209	55	385	27.156	---	---	---	6.19 av

and the estimating equation is

$$\log (y - 2.00) = 1.559 - 0.1522x$$

or

$$y = 10^{1.559-0.1542x} + 2.00.$$

The results are shown plotted on arithmetic grids in Fig. 35a, and $(y - \alpha)$ and y are plotted on semi-logarithmic grids in Fig. 35b. The extent to which the addition of the constant α improved the situation can be seen by comparing the average deviation of 6.19 calculated using the constant α with an average deviation of 10.2 calculated in the straight semi-log example. The same data were used in both cases.

THE POWER FUNCTION

The general equation of the power function is

$$y = ax^b.$$

As was true of the exponential, there is no direct least-squares solution for fitting the power function. Iterative methods can be used to achieve quite close approximations, but require such extensive calculation that they are only practical when a computer is available.* The usual practice is to transform the power function by taking the logs of both sides as

$$\log y = \log a + b \log x.$$

The result is a linear equation in terms of the logarithms of both x and y. When this transformation is reflected in the data by substituting the log of y for y and the log of x for x, the appropriate values for the example shown in Table 10 and Fig. 36a may be calculated and used in the normal equations for a straight line as follows:

$$\sum \log y = N \log a + b \sum \log x,$$

$$\sum \log x \log y = a \sum \log x + b \sum \log^2 x,$$

$$9.6245 = 10a + 10.8786b,$$

$$8.5378 = 10.8786a + 14.3460b,$$

$$\log a = 1.7994,$$

$$b = -0.7694.$$

*Graver and Boren, p. 75.

We must recognize that, as before, the result is a least-squares fit in terms of the logs rather than the actual values of y. The line will be placed such that the relative, not the absolute, deviations have been minimized.

Table 10

USING THE METHOD OF LEAST SQUARES TO FIT THE POWER FUNCTION: WORKSHEET

y	$\log y$	x	$\log x$	$\log^2 x$	$\log x \log y$	$\log y_c$	y_c	Percent Deviation
50	1.6990	2	0.3010	0.0906	0.5114	1.5678	36.98	26.0
25	1.3979	3	0.4771	0.2276	0.6669	1.4324	27.07	-8.3
20	1.3010	5	0.6990	0.4886	0.9094	1.2617	18.27	8.7
13	1.1139	6	0.7782	0.6056	0.8668	1.2007	15.88	-22.2
10	1.0000	10	1.0000	1.0000	1.0000	1.0300	10.72	-7.2
6	0.7782	15	1.1761	1.3832	0.9152	0.8946	7.84	-30.7
6	0.7782	20	1.3010	1.6926	1.0124	0.7984	6.29	-4.8
4	0.6021	40	1.6021	2.5667	0.9646	0.5668	3.69	-7.8
3	0.4771	50	1.6990	2.8866	0.8106	0.4923	3.11	-3.7
3	0.4771	70	1.8451	3.4044	0.8803	0.3798	2.40	20.0
140	9.6245	221	10.8786	14.3459	8.5377	---	---	13.94 av

The estimating equation expressed in logarithmic form is

$$\log y = 1.7994 - 0.7694 \log x.$$

The same equation expressed as a power function is

$$y = 63.01x^{-0.7694}.$$

As was the case with the exponential, the addition of a constant

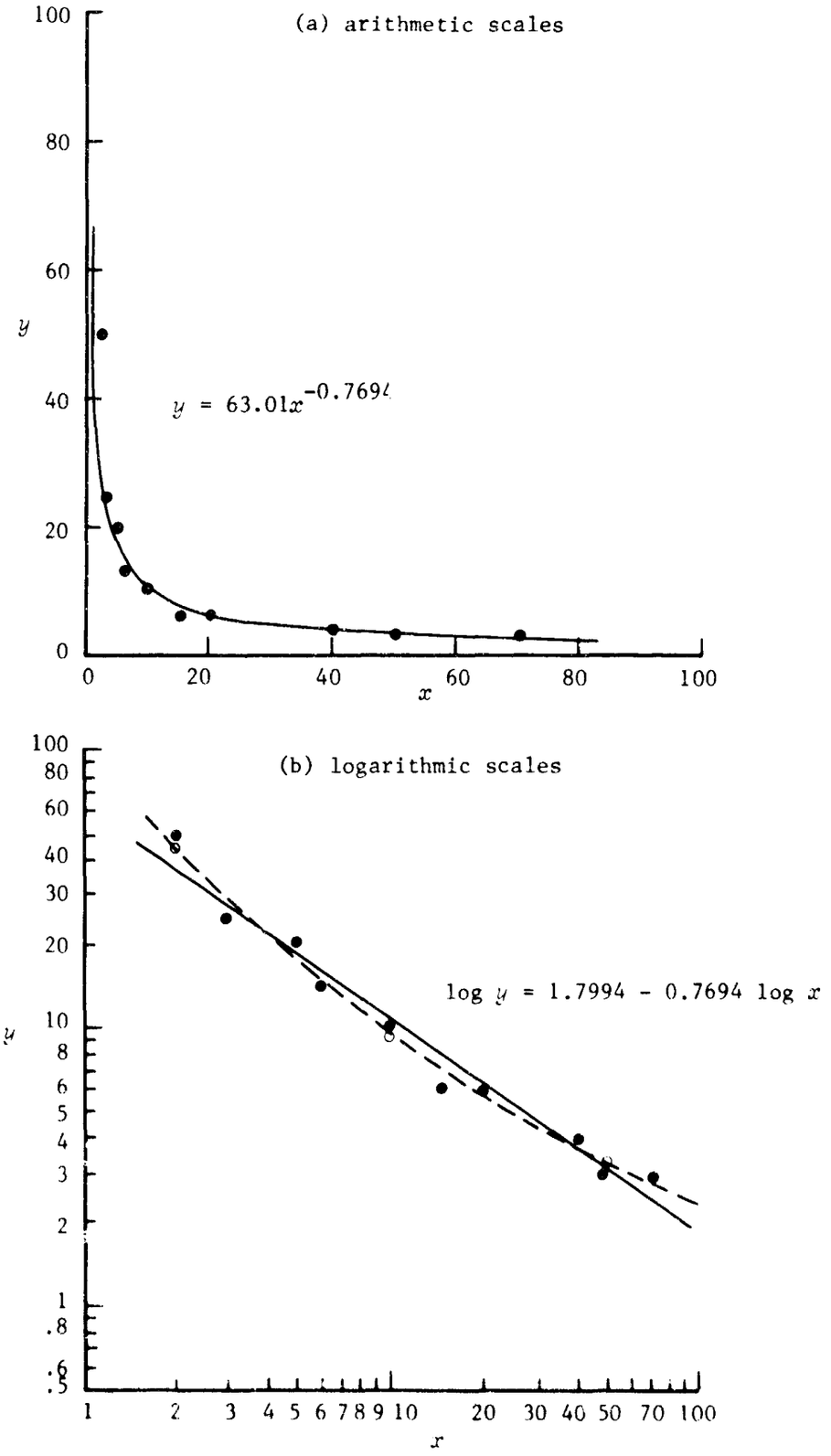


Fig. 36--Power function fitted using the method of least squares

to the equation

$$y = ax^b,$$

as in

$$y = ax^b + \alpha,$$

can often help in using the power function to describe a relationship as slightly curvilinear in terms of the logs of both x and y.

In log form the equation including the constant α is

$$\log (y - \alpha) = \log a + \log x.$$

Although the constant α can be determined here by trial and error, it is more conveniently estimated by much the same formula as for the exponential case:*

$$\alpha = \frac{y_1 y_2 - y_3^2}{y_1 + y_2 - 2y_3}.$$

In this case, it is easier to plot the data on logarithmic coordinate paper, and to draw the smooth curve as before. We select three points falling on the curve, two at the extremities and one in between such that its x coordinate is the geometric mean of the coordinates of the other two points, as

$$x_3 = \sqrt{x_1 x_2}.$$

The entire procedure is illustrated in Table 11 by Fig. 36(b), 37(a) and 37(b). The extent to which the addition of the constant α improved the result can be seen by comparing the average deviations.

---

*See Appendix B.

The calculation of α is as follows:*

$$P_1 = (x_1, y_1) = (2, 46),$$

$$P_2 = (x_2, y_2) = (50, 3.2),$$

$$x_3 = \sqrt{x_1 x_2} = \sqrt{100} = 10,$$

$$P_3 = (x_3, y_3) = (10, 9.5),$$

$$\alpha = \frac{y_1 y_2 - y_3^2}{y_1 + y_2 - 2y_3},$$

$$\alpha = \frac{(3.2)(46) - (9.5)(9.5)}{3.2 + 46 - 2(9.5)},$$

$$\alpha = 1.89.$$

The normal equations and their solution are given below:

$$\sum \log (y - \alpha) = Na + b \sum \log x,$$

$$\sum (\log x)[(\log (y - \alpha)] = a \sum \log x + b \sum (\log x)^2,$$

$$7.9013 = 10a + 10.8785b,$$

$$5.9598 = 10.8785a + 14.3457b,$$

$$a = 1.9318,$$

$$b = -1.0494.$$

The equation is

$$\log (y - \alpha) = 1.9318 - 1.0494 \log x,$$

*See example, p. 78.

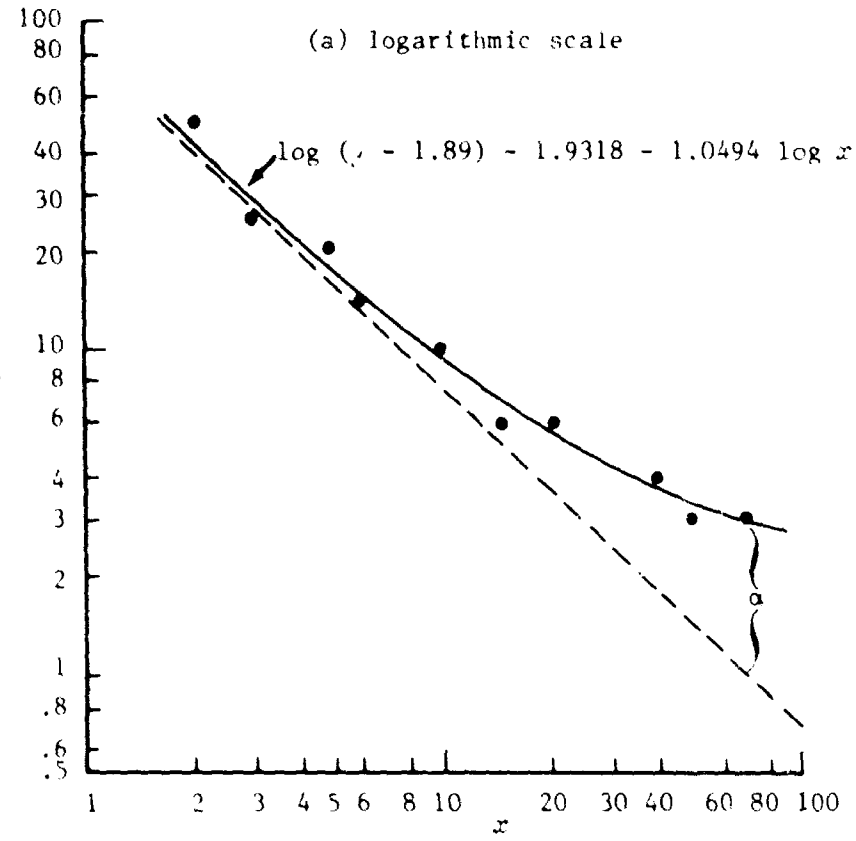
or as a power function,

$$y = 83.35x^{-1.0494} + 1.89.$$

Table 11

USING THE METHOD OF LEAST SQUARES TO FIT THE POWER FUNCTION WITH THE CONSTANT α: WORKSHEET

y	$(y - a)$	$\log(y - a)$	x	$\log x$	$(\log x)^2$	$(\log x)\lvert\log(y - a)\rvert$	$\log(y - a)_c$	$(y - a)_c$	y_c	Percent Deviation
50	48.11	1.6822	2	0.3010	0.0906	0.5064	1.6159	41.29	43.18	13.6
25	23.11	1.3638	3	0.4771	0.2276	0.6507	1.4311	26.98	28.87	-15.5
20	18.11	1.2579	5	0.6989	0.4886	0.8792	1.1982	15.78	17.67	11.7
13	11.11	1.0457	6	0.7782	0.6055	0.8137	1.1151	13.04	14.93	-14.9
10	8.11	0.9090	10	1.0000	1.0000	0.9090	0.8823	7.63	9.52	4.8
6	4.11	0.6138	15	1.1761	1.3832	0.7219	0.6975	4.98	6.87	-14.5
6	4.11	0.6138	20	1.3010	1.6927	0.7986	0.5664	3.68	5.57	7.2
4	2.11	0.3243	40	1.6021	2.5666	0.5195	0.2505	1.78	3.67	8.3
3	1.11	0.0453	50	1.6990	2.8865	0.0770	0.1488	1.41	3.30	-10.0
3	1.11	0.0453	70	1.8451	3.4044	0.0836	-0.0046	0.99	2.88	4.0
140	121.10	7.9013	221	10.8785	14.3457	5.9598	---	---	---	10.43 av



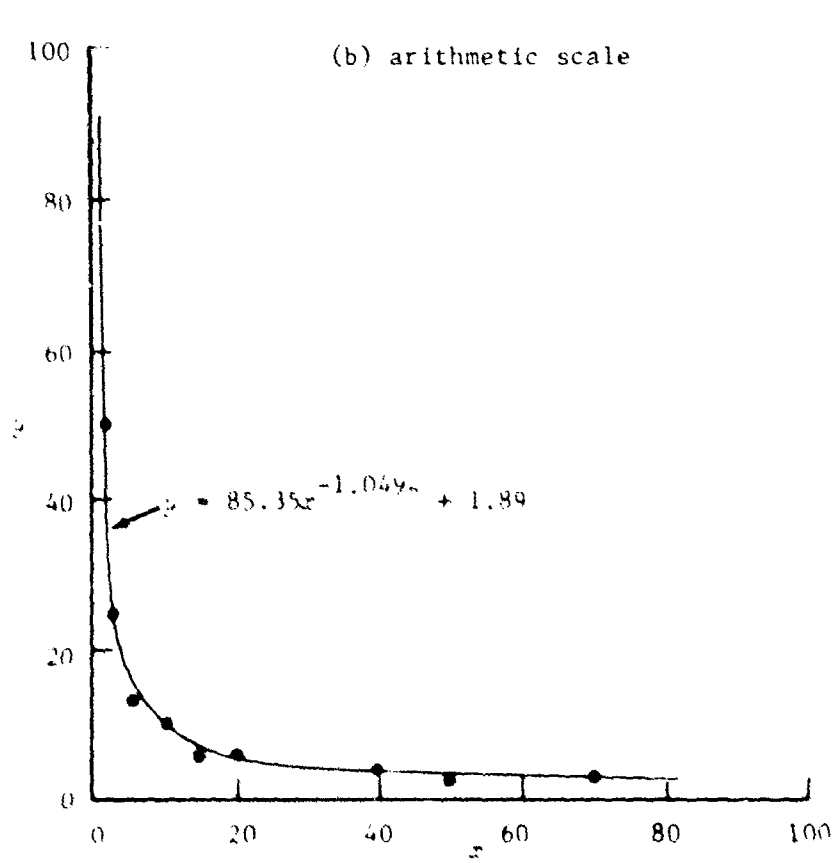


Fig. 37--Power function with constant c_1 fitted using the method of least squares

III. THREE-VARIABLE CURVE FITTING

THE LINEAR CASE

An empirical equation used to describe a three-variable linear relationship has the general form

$$x_1 = a + b_2 x_2 + b_3 x_3,$$

which is a simple extension of the two-variable linear (straight line) equation previously discussed.* To be consistent with the three-variable equation, we will write the two-variable relationship as:

$$x_1 = a + b_2 x_2.$$

We have already learned that the constant term, a, was the value of x_1 when x_2 was equal to 0. We further learned that b_2 was called the slope of the straight line and that, depending on whether b_2 was positive or negative, the value of b_2 determined the extent to which x_1 would be increased or decreased with changes in x_2.

The three-variable relationship may be thought of as two two-variable relationships interacting with each other. For example,

$$x_1 = a + b_2 x_2$$

and

$$x_1 = a + b_3 x_3$$

are two separate two-variable linear relationships, the first describing the impact of x_2 on the value of x_1 and the second the impact of x_3 on x_1. In the first relationship, however, the extent to which x_3

*When writing multi-variable equations, it is conventional to use the subscripted x_i as above rather than y and x as has been done previously.

influences the value of X_1 is not accounted for, nor, in the second, is the extent to which X_2 influences the value of X_1. What we really need is a relationship between X_1 and X_2 and between X_1 and X_3 where in each case the effect of the other independent variable on X_1 has been eliminated. Assuming that it is possible to obtain these, we write

$$X_1 = a_{1.23} + b_{12.3}X_2$$

and

$$X_1 = a_{1.23} + b_{13.2}X_3.$$

where subscripts indicate the variable whose effect has been eliminated. In the equation above, a is identified by the subscript 1.23, indicating that a is the value of X_1, once the effects of X_2 and X_3 have been eliminated. Since a is a constant, the relationship is a simple one; when X_2 and X_3 are eliminated from consideration, X_1 is in fact equal to $a_{1.23}$. In this equation the slope b is subscripted 12.3, indicating that $b_{12.3}$ is the net slope of the relationship between X_1 and X_2 independent of the impact of X_3. The numbers to the left of the decimal point in the subscript identify which two variables are being related; those to the right identify the variables whose effects have been eliminated. The subscripts used follow a logical pattern, and in fact this scheme of subscripting is often extended to four or more variable relationships.

Now, given that in each of the two straight line relationships shown above we have pure relationships (net relationships) between X_1 and each of the independent variables, and given that the two

independent variables completely determine the value of X_1, it is proper to combine them to write

$$X_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3;$$

this is the three-variable linear relationship with which we began. The coefficients of X_2 and X_3, $b_{12.3}$ and $b_{13.2}$, are frequently referred to as net regression coefficients and are in fact the slopes of the two separate straight lines described above. Each describes the impact of its accompanying variable on the dependent variable X_1. The constant $a_{1.23}$ is simply interpreted as the value of X_1 when both X_2 and X_3 are equal to 0.

To explore the idea of a net regression coefficient further and, at the same time, to illustrate one way that this type of relationship can be fitted to actual data, we will use the following example. In this case, we will begin with the answer and use a curve-fitting technique to see how closely we can reproduce it.

Assume that we are going to publish a technical report and we are concerned about the cost consequences of including various combinations of illustrations and plain printed pages. We contact a number of prospective printers and find that, on the average, for each report printed, there are three charges: a fixed charge of \$1.00; a charge of \$0.10 per illustration; and \$0.04 per printed page. The charges may be more concisely stated in the following three-variable linear relationship:

$$C = \$1.00 + \$0.10I + \$0.04P,$$

where ' = the cost per report,

' = the number of illustrations per report,

' = the number of printed pages per report.

At this point, we arbitrarily select a number of possibilities, choosing some with differing numbers of printed pages and a fixed number of illustrations and others with varying numbers of illustrations and the same number of printed pages. Further, for each combination chosen, we use the above cost equation to determine what it would cost to print the particular report. We select twelve reports as shown in Table 12, each with a different combination of illustrations and printed pages, and determine the printing cost of each.

Table 12

DATA ON SELECTED REPORTS

Report No.	No. of Illustrations (I)	No. of Printed Pages (P)	Cost to Print per Copy ($) ($C$)
1	1	18	1.82
2	2	4	1.36
3	2	10	1.60
4	2	20	2.00
5	3	15	1.90
6	4	13	1.92
7	5	7	1.78
8	5	16	2.14
9	6	6	1.84
10	6	2	1.68
11	7	1	1.74
12	7	7	1.98

Taking Report No. 3 as an example, we can see that the cost of $1.60 is arrived at as follows:

Fixed charge	$1.00
Illustrations (2) @ $0.10 . .	.20
Printed pages (10) @ $0.04 . .	.40
Total	$1.60

Let us now assume that, instead of having the equation which allowed us to calculate the costs above, we have only the data contained in Table 12 and we wish to find the equation. In such an example (which is unlike the usual case) we will assume that we know the price to be influenced only by the two variables, number of illustrations (I) and number of printed pages (P).

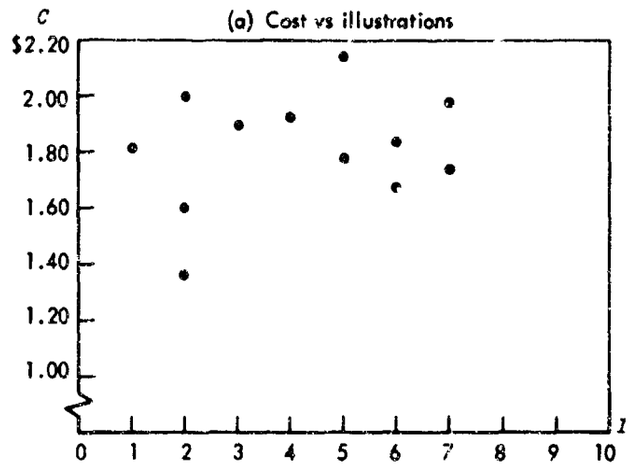
As has been our practice in the past in attacking such problems, we begin by constructing scatter diagrams, but, because it is difficult (although possible) to construct three-dimensional scatter diagrams,* we will be content with the more usual two-dimensional diagrams. In doing this, let us think in terms of the two two-variable straight lines discussed earlier. We begin by plotting the cost (C) against the number of illustrations (I) on one graph and the cost (C) against the number of printed pages (P) on the other. The first two diagrams (a and b) in Fig. 38 show the results. As we should have expected, in neither case do we see a clearly defined relationship. Any relationship that might exist between cost and the number of illustrations is obviously distorted by the fact that reports with the same number of illustrations have different numbers of printed pages.

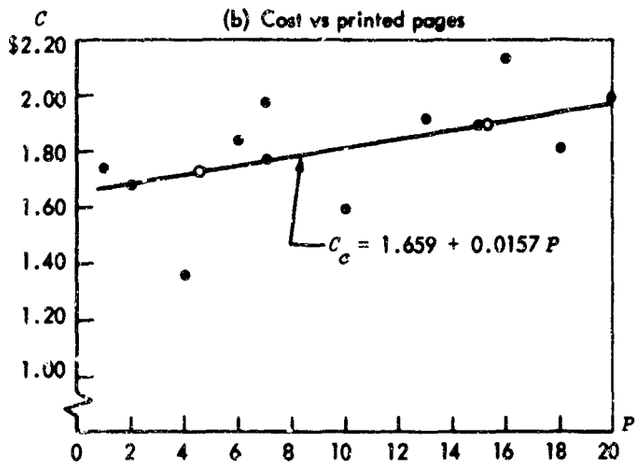
---

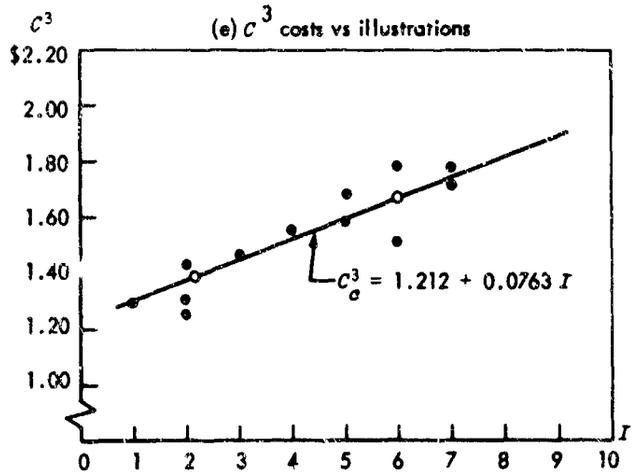
*W. A. Spurr and C. P. Bonini, *Statistical Analysis for Business Decisions*, Richard D. Irwin, Inc., Homewood, Illinois, 1967, p. 592.

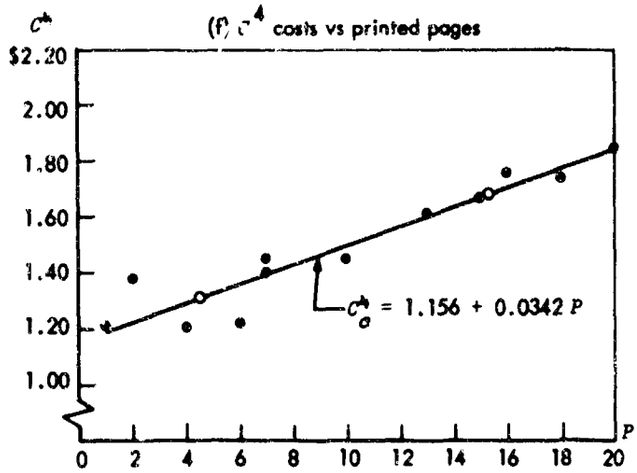
For example, there are three reports each with two illustrations, but one has four, one has ten, and one twenty printed pages. The number of illustrations similarly distorts the relationship between cost and number of printed pages shown in Fig. 38b. Even with all of the distortion present, it is possible to see a general upward trend in Fig. 38b. As the number of printed pages increases there is a commensurate increase in the cost. Our curve-fitting technique will be to capitalize on this by fitting a straight line to the data plotted in Fig. 38b and to use the results to improve the relationship between cost and the number of illustrations. For simplicity we will use the method of averages to fit the straight line and the point-slope formula to write the required equation. The details of these and other required computations are shown in Table 13. When using the method of averages, the data are first ordered according to the value of the independent variable (see Columns a, b, c, d) of Table 13. Because there are two independent variables involved and because the data cannot be ordered according to both of them at the same time, two separate setups are required. Those calculations that require ordering according to number of illustrations are shown on the upper half of Table 13, and those that require ordering according to number of printed pages are shown on the lower half of Table 13. Since the sequence requires stepping back and forth between the upper and the lower half, the steps are indicated by the numbers shown in circles at the head of each column.

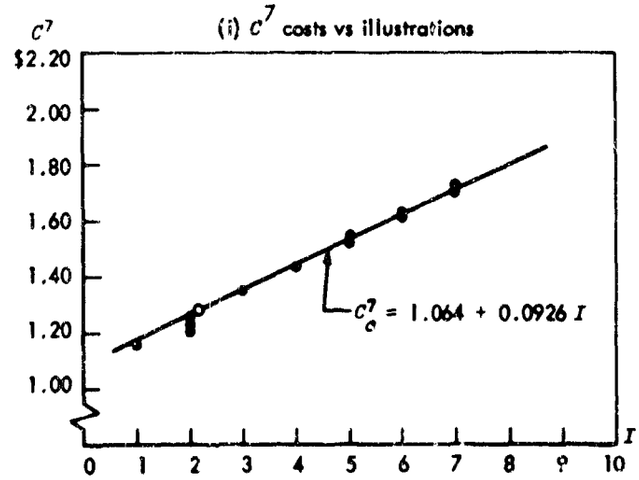
The calculations of the average points for fitting the first straight line (between cost and number of printed pages) are shown in the lower half of the table in Column 1. The coordinates of the two











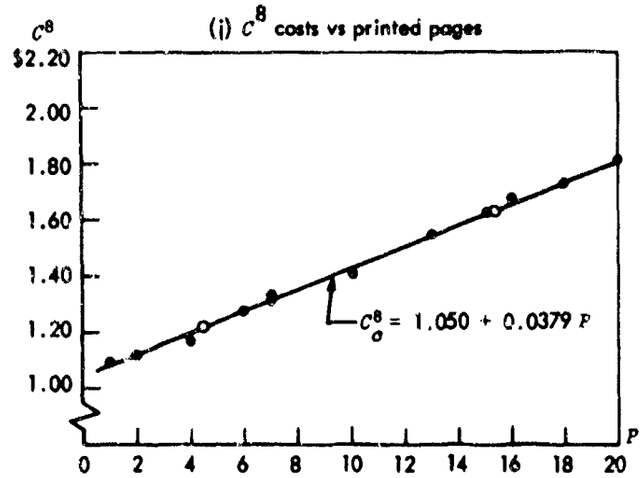
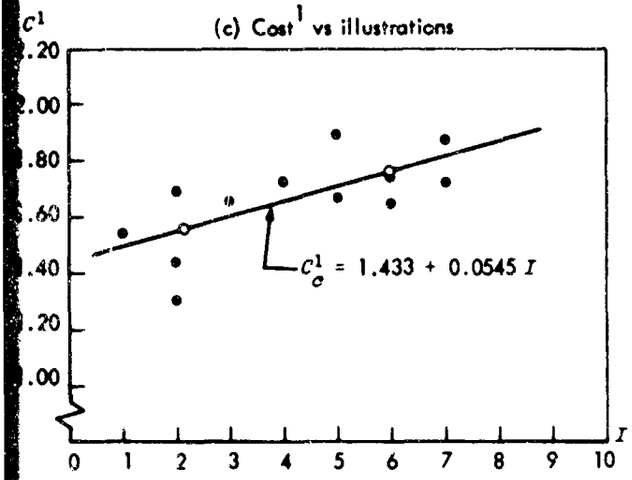
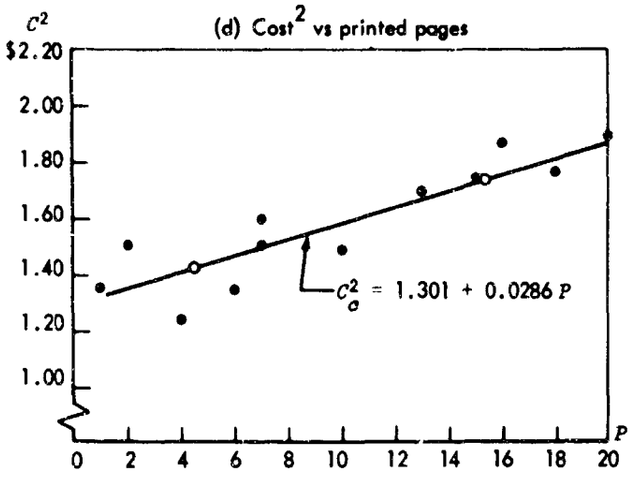
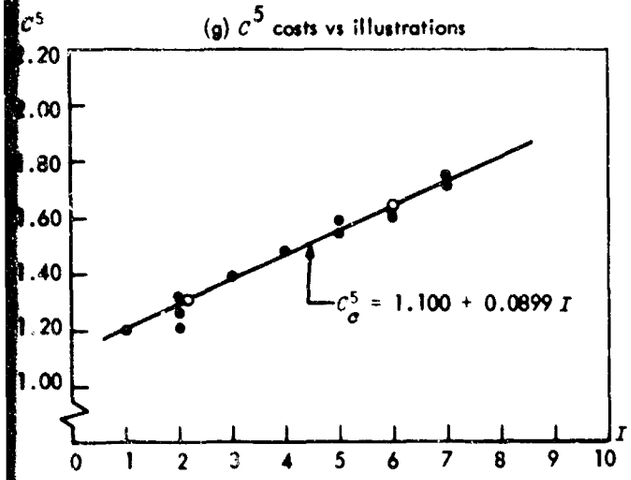
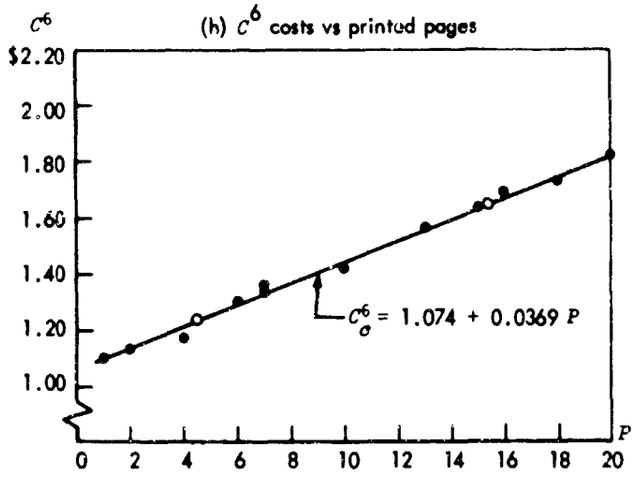


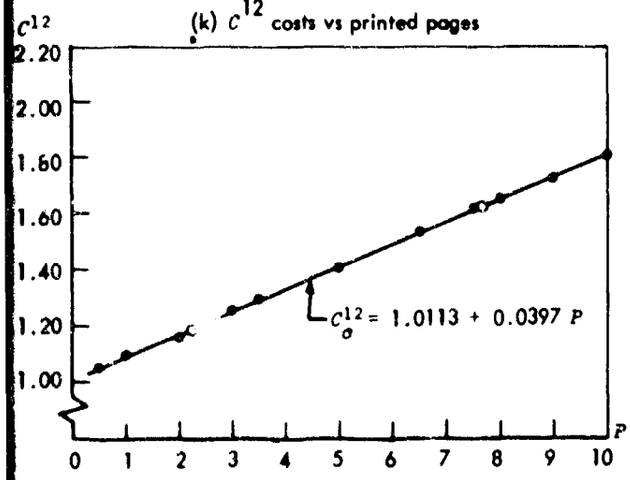
Fig. 38--Using successive approximations to ac

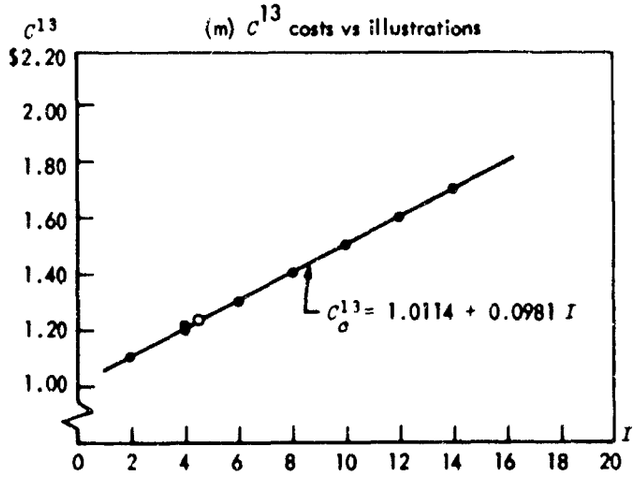












chieve continuously improved relationships

Table

USING SUCCESSIVE APPROXIMATIONS AND THE METHOD OF AVE

Expression: C =

Data									
(a) Report No.	(b) No. of Illustrations (I)	(c) No. of Printed Pages (P)	(d) Cost per Report (C)	③ (c^1)	④ (c^2)	⑦ (c^3)	⑧ (c^4)	⑪ (c^5)	
			(Fig. 38a)	(Fig. 38c)		(Fig. 38e)		(Fig. 38g)	
1	1	18	1.82	1.54	1.77	1.31	1.74	1.20	
2	2	4	1.36	1.30	1.25	1.25	1.21	1.22	
3	2	10	1.60	1.44	1.49	1.31	1.45	1.26	
4	2	20	2.00	1.69	1.89	1.43	1.85	1.32	
5	3	15	1.90	1.66	1.74	1.47	1.67	1.39	
6	4	13	1.92	1.72	1.70	1.55	1.61	1.48	
	Av. = 2.33			Av. = 1.56		Av. = 1.39		Av. = 1.31	
7	5	7	1.78	1.67	1.51	1.58	1.40	1.54	
8	5	16	2.14	1.89	1.87	1.68	1.76	1.59	
9	6	2	1.68	1.65	1.35	1.62	1.22	1.61	
10	6	6	1.84	1.75	1.51	1.67	1.38	1.63	
11	7	1	1.74	1.72	1.36	1.71	1.21	1.71	
12	7	7	1.98	1.87	1.60	1.78	1.45	1.74	
	Av. = 6.00			Av. = 1.76		Av. = 1.67		Av. = 1.64	
	Δ = 3.67			Δ = 0.20		Δ = 0.28		Δ = 0.33	
				$c^1 = 1.433 + 0.0545I$	$c^2 = c - 0.0545I$	$c^3 = 1.212 + 0.0763I$	$c^4 = c - 0.0763I$	$c^5 = 1.100 + 0.0899I$	
			① (C)	② (c^1)	⑤ (c^2)	⑥ (c^3)	⑨ (c^4)	⑩ (c^5)	
			(Fig. 38b)		(Fig. 38d)		(Fig. 38f)		
11	7	1	1.74	1.72	1.36	1.71	1.21	1.71	
9	6	2	1.68	1.65	1.35	1.62	1.22	1.61	
2	2	4	1.36	1.30	1.25	1.25	1.21	1.22	
10	6	6	1.84	1.75	1.51	1.67	1.38	1.63	
7	5	7	1.78	1.67	1.51	1.58	1.40	1.54	
12	7	7	1.98	1.87	1.60	1.78	1.45	1.74	
		Av. = 4.50	Av. = 1.73		Av. = 1.43		Av. = 1.31		
3	2	10	1.60	1.44	1.49	1.31	1.45	1.26	
6	4	13	1.92	1.72	1.70	1.55	1.61	1.48	
5	3	15	1.90	1.66	1.74	1.47	1.67	1.39	
8	5	16	2.14	1.89	1.87	1.68	1.76	1.59	
1	1	18	1.82	1.54	1.77	1.31	1.74	1.20	
4	2	20	2.00	1.69	1.89	1.43	1.85	1.32	
		Av. = 15.33	Av. = 1.90		Av. = 1.74		Av. = 1.68		
		Δ = 10.83	Δ = 0.17		Δ = 0.31		Δ = 0.37		
			$c = 1.659 + 0.0157P$	$c^1 = c - 0.0157P$	$c^2 = 1.301 + 0.0286P$	$c^3 = c - 0.0286P$	$c^4 = 1.156 + 0.0342P$	$c^5 = c - 0.0342P$	

Table 13

OD OF AVERAGES TO FIT A THREE-VARIABLE LINEAR EQUATION: WORKSHEET

on: $C = 1.00 + 0.10I + 0.04P$

Calculations

⑪ (c^5)	⑫ (c^6)	⑬ (c^7)	⑭ (c^8)	⑮ (c^9)	⑳ (c^{10})	㉓ (c^{11})	㉔ (c^{12})	㉗ (c^{13})
g. 38g)		(Fig. 38i)						(Fig. 38m)
1.20	1.73	1.16	1.73	1.14	1.72	1.12	1.72	1.11
1.22	1.18	1.21	1.17	1.20	1.17	1.20	1.16	1.20
1.26	1.42	1.23	1.41	1.22	1.41	1.21	1.40	1.20
1.32	1.82	1.26	1.81	1.24	1.81	1.22	1.80	1.21
1.39	1.63	1.35	1.62	1.33	1.61	1.32	1.61	1.30
1.48	1.56	1.44	1.55	1.43	1.54	1.42	1.53	1.40
= 1.31		Av. = 1.28		Av. = 1.26		Av. = 1.25		Av. = 1.24
1.54	1.33	1.52	1.32	1.51	1.30	1.51	1.29	1.50
1.59	1.69	1.55	1.68	1.53	1.66	1.52	1.65	1.50
1.61	1.14	1.61	1.12	1.60	1.11	1.60	1.09	1.60
1.63	1.30	1.62	1.28	1.61	1.27	1.61	1.25	1.60
1.71	1.11	1.70	1.09	1.70	1.07	1.70	1.05	1.70
1.74	1.35	1.72	1.33	1.71	1.31	1.71	1.29	1.70
= 1.64		Av. = 1.62		Av. = 1.61		Av. = 1.61		Av. = 1.60
= 0.33		Δ = 0.34		Δ = 0.35		Δ = 0.36		Δ = 0.36
00+0.0899I	$c^6=C-0.0899I$	$c_o^7=1.064+0.0926I$	$c^8=C-0.0926I$	$c_o^9=1.035+0.0954I$	$c^{10}=C-0.0954I$	$c_o^{11}=1.021+0.0981I$	$c^{12}=C-0.0981I$	$c_o^{13}=1.0114+0.0981I$

⑩ (c^5)	⑪ (c^6)	⑭ (c^7)	⑬ (c^8)	⑱ (c^9)	㉑ (c^{10})	㉒ (c^{11})	㉕ (c^{12})	㉙ (c^{13})
	(Fig. 38h)		(Fig. 38j)				(Fig. 38k)	
1.71	1.11	1.70	1.09	1.70	1.07	1.70	1.05	1.70
1.61	1.14	1.61	1.12	1.60	1.11	1.60	1.09	1.60
1.22	1.18	1.21	1.17	1.20	1.17	1.20	1.16	1.20
1.63	1.30	1.62	1.28	1.61	1.27	1.61	1.25	1.60
1.54	1.33	1.52	1.32	1.51	1.30	1.51	1.25	1.50
1.74	1.35	1.72	1.33	1.71	1.31	1.71	1.29	1.70
	Av. = 1.24		Av. = 1.22		Av. = 1.21		Av. = 1.19	
1.26	1.42	1.23	1.41	1.22	1.41	1.21	1.40	1.20
1.48	1.56	1.44	1.55	1.43	1.54	1.42	1.53	1.40
1.39	1.63	1.35	1.62	1.33	1.61	1.32	1.61	1.30
1.59	1.69	1.55	1.68	1.53	1.66	1.52	1.65	1.50
1.20	1.73	1.16	1.73	1.14	1.72	1.12	1.72	1.11
1.32	1.82	1.26	1.81	1.24	1.81	1.22	1.80	1.21
	Av. = 1.64		Av. = 1.63		Av. = 1.63		Av. = 1.62	
	Δ = 0.40		Δ = 0.41		Δ = 0.42		Δ = 0.43	
C-0.0342P	$c_o^6=1.074+0.0369P$	$c^7=C-0.0369P$	$c_o^8=1.030+0.0379P$	$c^9=C-0.0379P$	$c_o^{10}=1.035+0.0388P$	$c^{11}=C-0.0388P$	$c_o^{12}=1.0113+0.0397P$	$c^{13}=C-0.0397P$

points (P_1, C_1) and (P_2, C_2) are (4.50, 1.73) and (15.33, 1.90) respectively. The modified point-slope formula is

$$\quad - \quad_1 = \left(\frac{\quad_2 - \quad_1}{\quad_2 - \quad_1}\right)(\quad - \quad_1).$$

To simplify calculation with a desk calculator, the modified point-slope formula above was recast as follows:

$$\quad = \frac{\quad_1 \quad_2 - \quad_2 \quad_1}{\quad_2 - \quad_1} + \left(\frac{\quad_2 - \quad_1}{\quad_2 - \quad_1}\right)\quad .$$

or

$$\quad = a + b\quad ,$$

where

$$a = \frac{\quad_1 \quad_2 - \quad_2 \quad_1}{\quad_2 - \quad_1},$$

$$b = \left(\frac{\quad_2 - \quad_1}{\quad_2 - \quad_1}\right).$$

In the first case, the values are substituted and

$$a = \frac{(1.73)(15.33) - (1.90)(4.50)}{15.33 - 4.50} = 1.659$$

and

$$b = \frac{1.90 - 1.73}{15.33 - 4.50} = 0.0157.$$

Because $\quad_2 - \quad_1$ and $\quad_2 - \quad_1$ are most easily calculated as the numbers are entered in the table, they should be done at that time and indicated by ' entered in the appropriate column.

The equation of the straight line describing the relationship between cost and number of printed pages is thus

$$C_p = 1.659 + 0.0157P.$$

The subscript p is used to indicate that values of C calculated from this equation are estimates rather than actuals. When this equation is plotted as in Fig. 38b, it gives a rough approximation of the true relationship. However, a rough approximation is better than none, as we shall subsequently see.

At the moment, the value of the constant a is of no interest. The value b or 0.0157 means that for each printed page we must add 1.57 cents to the cost. We can reduce the cost of each case by this figure in proportion to the number of printed pages, and then examine these results with respect to the number of illustrations. The adjustment is made by setting

$$C_1 = C - 0.0157P.$$

For report No. 11 the result would be

$$C_1 = 1.74 - (0.0157)(1),$$

$$C_1 = 1.72,$$

as shown in column 2 in the lower half of Table 13. We next make the same reduction in cost for each report in proportion to the number of printed pages. When this has been completed, we transfer the results to Column 3, in the upper portion of the table, and simultaneously

reorder them according to the number of illustrations in each case. We indicate these values by the symbol c^1 where 1 (known as a superscript, not an exponent) signifies the first adjustment to the original costs.

When we have plotted these adjusted costs against the number of illustrations as in Fig. 38c, we have a more definite relationship than that indicated in Fig. 38a. What has happened is this: Although the equation relating cost to number of printed pages was extremely rough, it was sufficient to eliminate enough of the effect of printed pages from c to clear up the relationship between c and I.

The next step is to follow our logic and determine the relationship between c^1 and I using the results to further clean up the relationship between cost and number of printed pages. Once again, we employ the method of averages, placing the results in Column 3 in the upper portion of Table 13. This fitted line can be seen plotted in Fig. 38c. The equation of the fitted line is

$$c^1 = 1.433 + 0.0545I.$$

This equation gives us an approximation of the impact of the number of illustrations on cost--in this case 5.45 cents per illustration. The costs are again adjusted as in Column 4 in the upper portion of Table 13, this time to eliminate the effect of the number of illustrations according to the approximation given above. This adjustment is made according to the formula

$$c^2 = c - 0.0545I,$$

where ^{2} indicates that the cost has been adjusted for the second time.

The adjusted figures are next transferred to Column 5, lower half of Table 13, and the results plotted against the number of printed pages as in Fig. 38d. A comparison of Fig. 38d with Fig. 38b shows the extent to which our first approximation of the cost of illustrations has improved the relationship between total cost and number of printed pages. This process of refining the approximations is continued first with respect to one of the independent variables and then the other. Each time an approximate relationship is obtained it is used to further adjust the cost; the adjusted cost is then related to the other independent variable and the process repeated again. The calculations in Table follow the adjustment process through thirteen times. The calculations of Columns 3 through 8 and Columns 12 and 13 in Table 13 are illustrated by Fig. 38d through Fig. 38m.

The relationship between cost and number of printed pages shown in Fig. 38k which was arrived at on the 12th adjustment can be described by the linear equation

$$C_c^{12} = 1.0113 + 0.0397P.$$

This equation is quite close to that portion of the original equation dealing with printed pages,

$$C = 1.00 + 0.04P.$$

The relationship between cost and number of illustrations shown in Fig. 38m is also quite close to the relevant part of the original equation:

$$_{13} = 1.0144 + 0.0981_{}$$

as compared to

$$' = 1.00 + 0.10_{}.$$

When the two two-variable equations are combined as

$$_{} = 1.01 + 0.098_{} + 0.039_{},$$

we have a very close representation of the original equation

$$' = 1.00 + 0.10_{} + 0.04_{}.$$

Had we continued with our process of successive approximation and adjustment, we could conceivably have reproduced the original equation exactly. But this would have meant carrying the calculations to more significant digits which was unnecessary for the purposes of this example. This method of curve fitting, quite appropriately called the Method of Successive Approximations, can be used quite generally--even in cases where the separate relationships can only be described by freehand non-mathematically describable curves.

Fortunately the method of least squares accomplishes similar results for the three-variable linear relationship by means of a direct and absolute rather than an approximate solution. To show that both methods result in the same solution, the method of least squares is next applied to the same problem. Data are calculated in Table 14, and the accompanying graphs plotted in Fig. 39. Normal equations for this solution are as follows:

$$\sum C = Na + b_1 \sum I + b_2 \sum P,$$

$$\sum IC = a \sum I + b_1 \sum I^2 + b_2 \sum IP,$$

$$\sum PC = a \sum P + b_1 \sum IP + b_2 \sum P^2.$$

This works out to

$$21.76 = 12a + 50b_1 + 119b_2,$$

$$91.88 = 50a + 258b_1 + 402b_2,$$

$$224.36 = 119a + 402b_1 + 1629b_2.$$

Therefore the solution is

$$C_c = 1.000 + 0.10I + 0.04P.$$

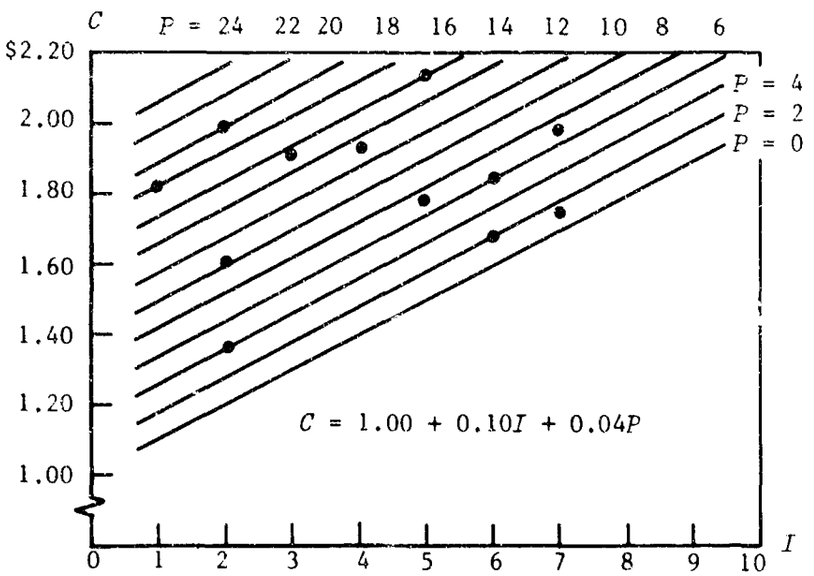


Fig. 39--Least-squares solution to the three-variable problem showing one way to graph a three-variable equation

Table 14

USING THE METHOD OF LEAST SQUARES TO FIT A THREE-VARIABLE LINEAR EQUATION: WORKSHEET

($C = 1.000 + 0.10I + 0.04P$)

Report No.	Cost per Report (C)	Illustrations (I)	Printed Pages (P)	$(I)(P)$	(I^2)	(P^2)	$(I)(C)$	$(P)(C)$
1	1.82	1	18	18	1	324	1.82	32.76
2	1.36	2	4	8	4	16	2.72	5.44
3	1.60	2	10	20	4	100	3.20	16.00
4	2.00	2	20	40	4	400	4.00	40.00
5	1.90	3	15	45	9	225	5.70	28.50
6	1.92	4	13	52	16	169	7.68	24.96
7	1.78	5	7	35	25	49	8.90	12.46
8	2.14	5	16	80	25	256	10.70	34.24
9	1.68	6	2	12	36	4	10.08	3.36
10	1.84	6	6	36	36	36	11.04	11.04
11	1.74	7	1	7	49	1	12.18	1.74
12	1.98	7	7	49	49	49	13.86	13.86
Total	21.76	50	119	402	258	1629	91.88	224.36

THE NONLINEAR CASE

It is not unusual to encounter sets of three or more variables that cannot be adequately described using linear relationships, and that require nonlinear curve fitting. In this section of the Memorandum we will use the method of least squares to fit the straight line, the exponential, the power function, and the parabola to a set of one dependent and two independent variables.

Fitting a three-variable linear equation and using the method of least squares has already been described. We remember that the linear equation

$$X_1 = a + b_2X_2 + b_3X_3$$

resulted from the two two-variable equations

$$X_1 = a + b_2X_2,$$

and

$$X_1 = a + b_3X_3,$$

with each describing the relationship between the dependent variable X_1 and either X_2 or X_3. In each case, the influence of the other was not accounted for. In the combined relationship, b_2 and b_3 were written $b_{12.3}$ and $b_{13.2}$ to show that in the first case the effect of X_3 was eliminated, and that in the second case the effect of X_2 was eliminated. The method of successive approximations was used to demonstrate how this could be done. Further, it was shown that the method of least squares produces the same answer with considerably less effort.

We will now build on these fundamentals to illustrate three-variable nonlinear curve fitting.

As there is nothing about the detailed calculations required here that is different from those previously illustrated, we will not describe them again. Instead, we will concentrate on showing how various nonlinear functional forms can be used. In particular, we will point up their peculiarities and consequently their limitations.

Twenty sets of the three variables--X_1, X_2, and X_3--are shown in Table 15. X_1 is the dependent variable; X_2 and X_3 are the independent variables. We will proceed to fit a linear, an exponential, a power function, and a parabolic relationship to these variables. Good practice dictates that we start by examining the data more closely. As with the two-variable case, preparing a scatter diagram is always a good beginning.

Figure 40 shows the results of plotting X_1 against X_2 while ignoring X_3. Little more than a general scattering of points is observed. But when each point is identified with its X_3 value and contour lines connecting all points with equal values of X_3 are drawn, as in Fig. 41, a relationship can be seen. For each value of X_3 X_1 increases with increases in X_2.

Figure 42 shows similar results. Here X_1 is plotted against X_3 and contours connecting points having equal values of X_2 have been drawn. For fixed values of X_2, X_1 increases with increases in X_3. At this point, we also note a distinct curvature in one or two of the contours which suggests a nonlinear relationship between X_1 and X_3.

A point from which to compare the results of fitting nonlinear relationships has been provided by fitting a linear relationship to

Table 15

RESULTS OF FITTING A THREE-VARIABLE LINEAR RELATIONSHIP
($X_1 = -20.01 + 0.4998X_2 + 1.295X_3$)

Observation	X_1	X_2	X_3	X_1(calc.)	X_1-X_1(calc.)	Percent Deviation
1	7.31	5	5	-11.08	18.39	251.5
2	37.67	5	49	45.92	-8.25	-22.0
3	67.37	5	71	74.42	-7.05	-10.5
4	121.31	5	100	111.99	9.38	-7.7
5	20.93	16	27	22.92	-1.99	-9.5
6	24.77	27	27	28.42	-3.65	-14.7
7	33.57	27	38	42.67	-9.10	-27.1
8	22.78	38	16	19.66	3.12	13.7
9	29.16	38	27	33.91	-4.75	-16.3
10	118.26	38	93	119.41	-1.15	-1.0
11	39.62	60	27	44.91	-5.29	-13.4
12	45.68	71	27	50.41	-4.73	-10.4
13	149.34	71	100	144.97	4.37	2.9
14	41.97	82	5	27.40	14.57	34.7
15	59.48	93	27	61.40	-1.92	-3.2
16	148.58	93	93	146.90	1.68	1.1
17	163.14	93	100	155.96	7.18	4.4
18	73.14	100	38	79.15	-6.01	-8.2
19	114.06	100	71	121.90	-7.84	-6.9
20	153.44	100	93	150.40	3.04	2.0
						23.0 av

the data. The least-squares normal equations and the resulting linear relationship follow:

$$\sum X_1 = Na + b_2 \sum X_2 + b_3 \sum X_3,$$

$$\sum X_1X_2 = a \sum X_2 + b_2 \sum X_2^2 + b_3 \sum X_2X_3,$$

$$\sum X_1X_3 = a \sum X_3 + b_2 \sum X_2X_3 + b_3 \sum X_3^2,$$

$$X_1 = -20.01 + 0.4998X_2 + 1.295X_3.$$

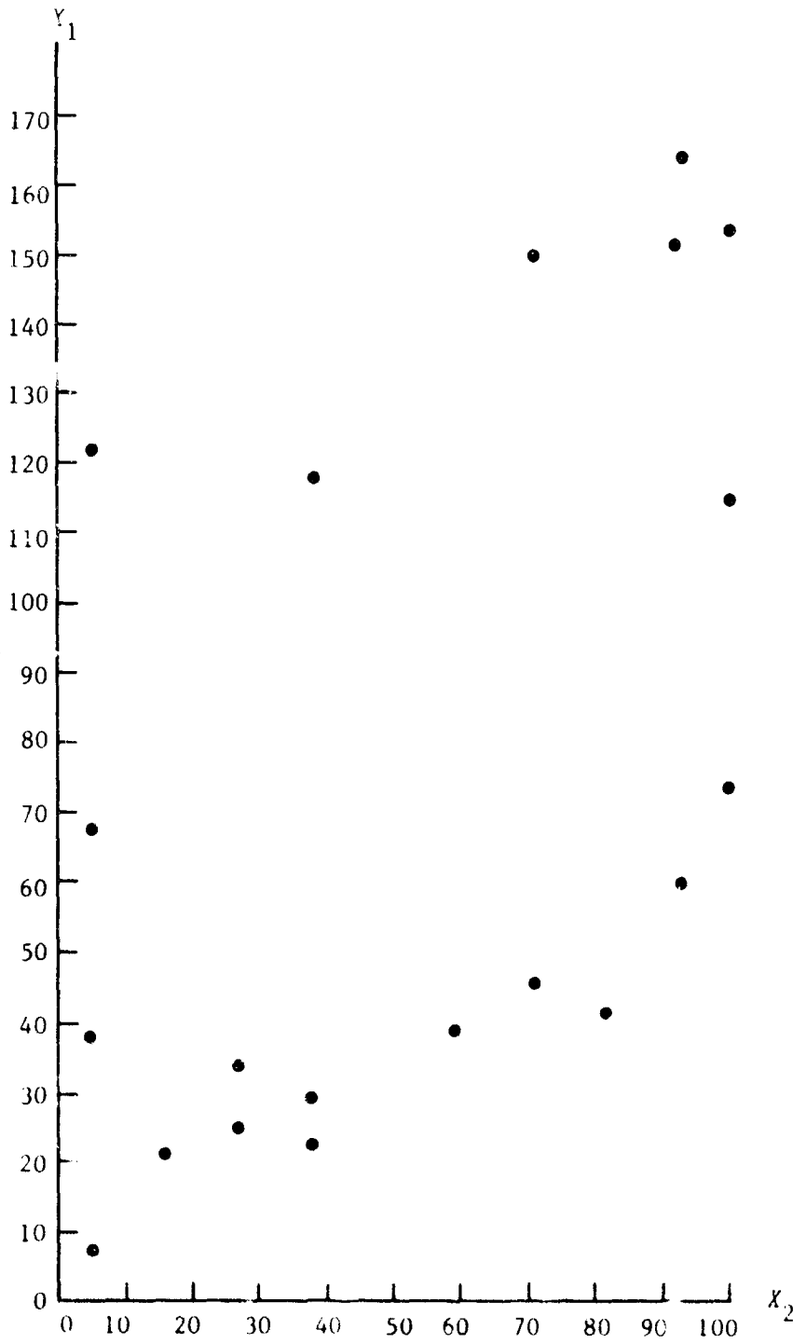


Fig. 40--Scatter diagram: X_1 vs X_2

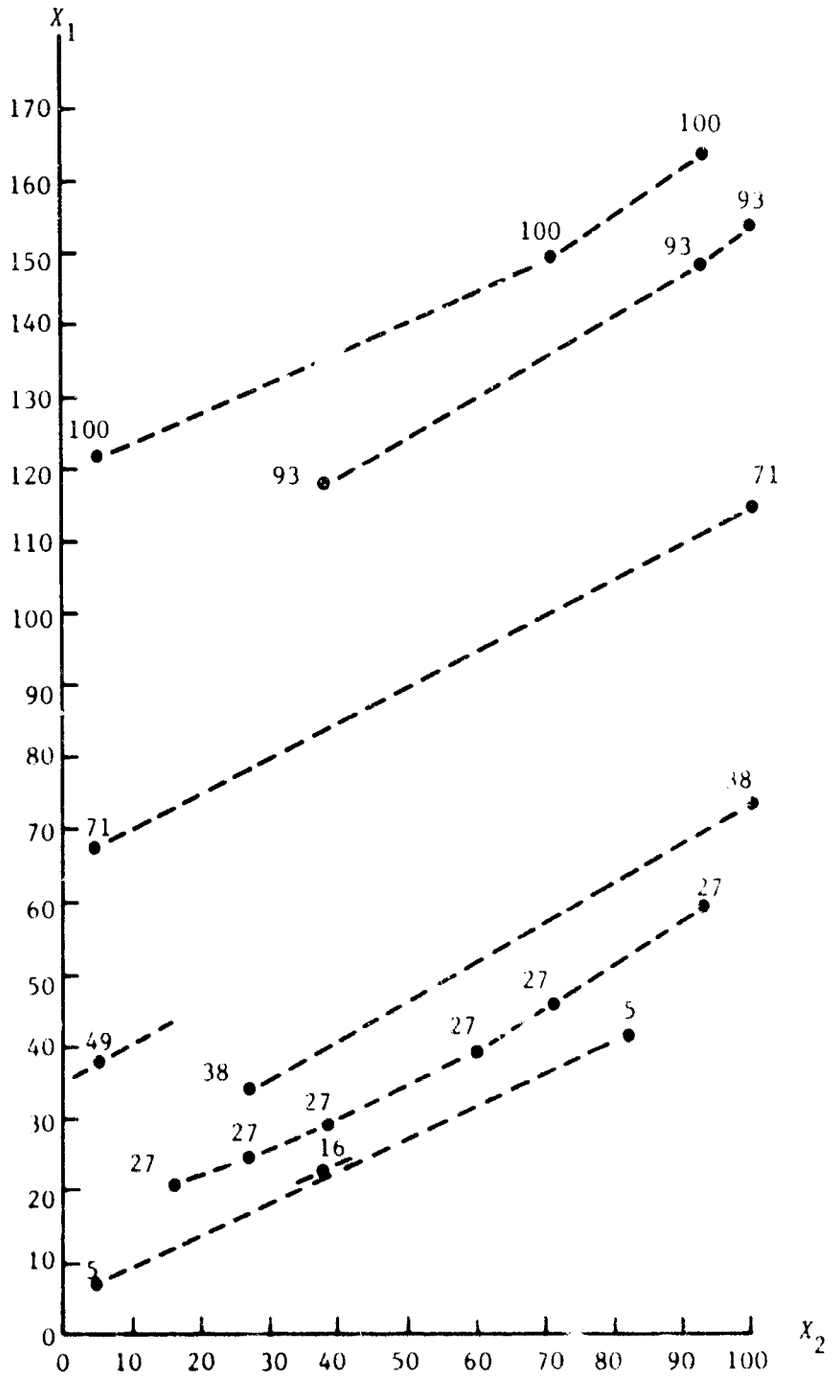


Fig. 41--Scatter diagram: X_1 vs X_2 with contours showing equal values of X_3

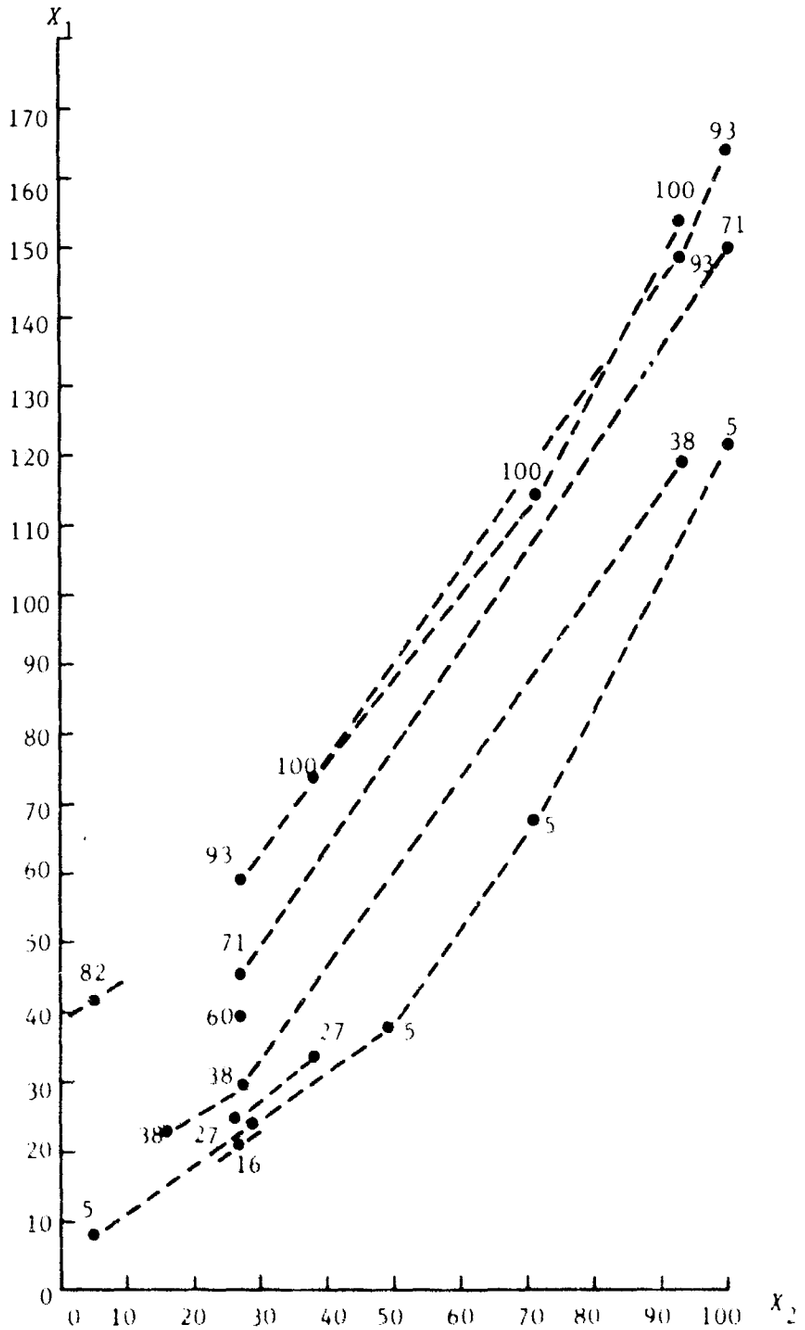


Fig. 42--Scatter diagram: X_1 vs X_3 with contours showing equal values of X_2

In Fig. 43, Y_1 is plotted against X_2 ignoring X_3. The straight lines result from solving the linear equation above, allowing X_2 to vary over its relevant range while holding X_3 constant at the values indicated in Fig. 43. The deviations of the points from the appropriate lines are indicated by the vertical connecting lines. As was to be expected, the linear relationship does not describe the data very well. A tabular presentation of the results was shown in Table 15.

Given the indications of nonlinearity in Figs. 41 and 42 and the poorness of fit achieved with the linear form, a nonlinear form seems in order. When confronted with a similar situation, analysts often turn immediately to the power function on the grounds that it will straighten anything out. We will try this and see what happens.

The basic power function in two variables is

$$Y_1 = aX_2^{b_2},$$

or

$$Y_1 = aX_3^{b_3}.$$

In logarithmic form these equations become

$$\log Y_1 = \log a + b_2 \log X_2,$$

and

$$\log Y_1 = \log a + b_3 \log X_3.$$

The transition from the two-variable equations to the one three-variable equation is analogous to the equations presented in the beginning of Section III.*

---

*See pp. 91-92.

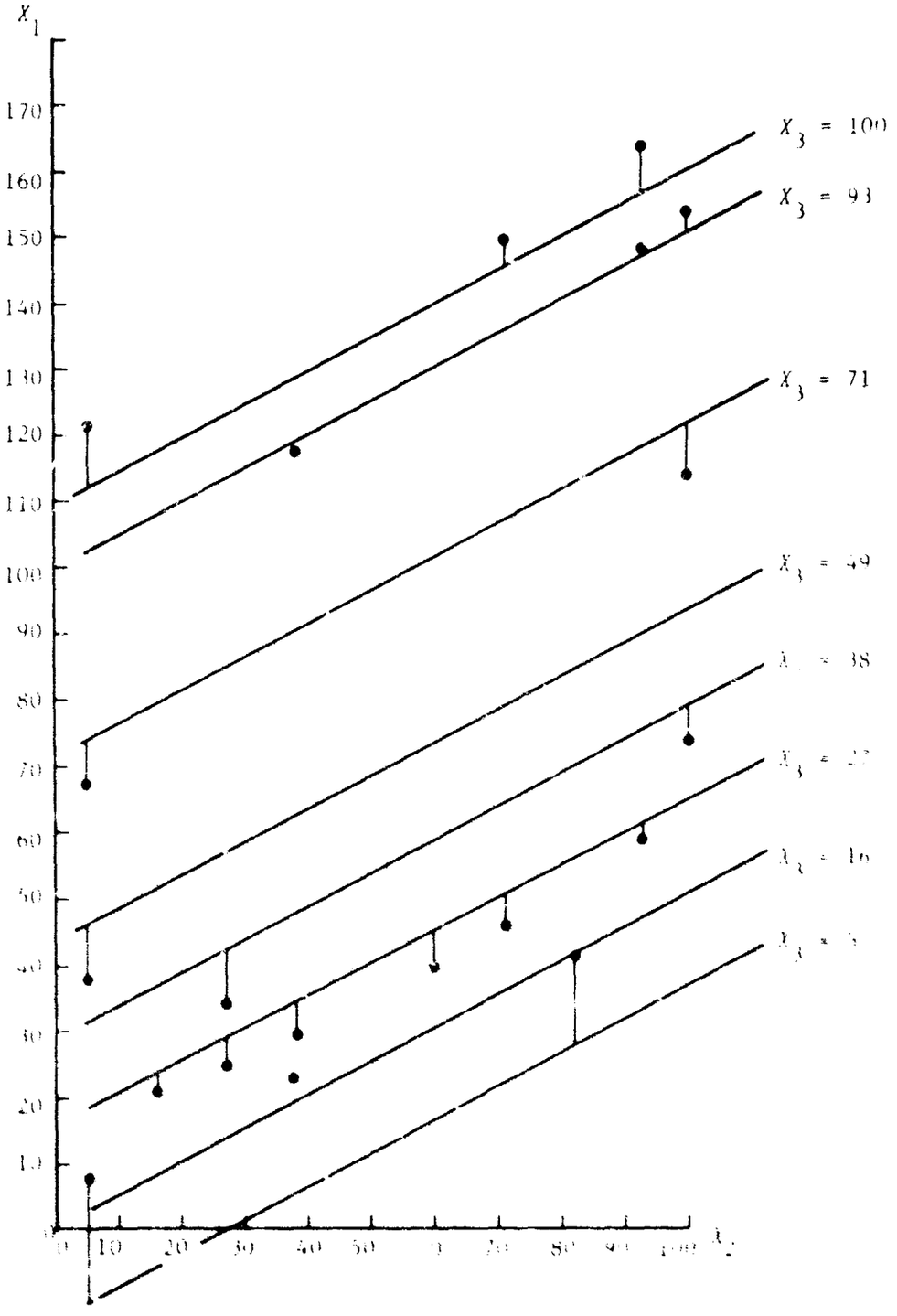


Fig. 43--Results of fitting a three-variable linear relationship (X_1 = -20.01 + 0.4998X_2 + 1.095X_3)

Either

$$\log X_1 = a + b_2 \log X_2 + b_3 \log X_3$$

or

$$X_1 = a X_2^{b_2} X_3^{b_3}$$

is the required equation and, as can be seen, the equation is linear in terms of the logarithms of the variables. The least-squares normal equations used before are appropriate here, given that the logarithms of the variables are substituted for the variables. For example,

$$\sum \log X_1 = Na + b_2 \sum \log X_2 + b_3 \sum \log X_3,$$

$$\sum \log X_1 \log X_2 = a \sum \log X_2 + b_2 \sum (\log X_2)^2 + b_3 \sum \log X_2 \log X_3,$$

$$\sum \log X_1 \log X_3 = a \sum \log X_3 + b_2 \sum \log X_2 \log X_3 + b_3 \sum (\log X_3)^2.$$

When the required values are calculated and this set of equations solved, the following power function results:

$$\log X_1 = 0.16555 + 0.26963 \log X_2 + 0.73198 \log X_3,$$

or

$$X_1 = 1.464 X_2^{0.26963} X_3^{0.73198}.$$

How well this equation does the job is shown in Fig. 44 and Table 16. It is obviously no better than the linear relationship and possibly even a little worse. The most striking shortcoming is that the direction of curvature is wrong. Figures 41 and 42 indicate that the required curve should be concave upwards, and these curves are concave

Table 16

RESULTS OF FITTING A THREE-VARIABLE POWER FUNCTION RELATIONSHIP
($\log X_1 = 0.16555 + 0.26963 \log X_2 + 0.73198 \log X_3$)

Observation	X_1	X_2	X_3	X (calc.)	X_1-X_1(calc.)	Percent Deviation
1	7.31	5	5	7.34	-0.03	-0.4
2	37.67	5	49	39.01	-1.34	-3.6
3	67.37	5	71	51.18	16.19	24.0
4	121.31	5	100	65.76	55.55	45.8
5	20.93	16	27	34.51	-13.58	-64.9
6	24.77	27	27	39.74	-14.97	-60.4
7	33.57	27	38	51.03	-17.46	-52.0
8	22.78	38	16	29.71	-6.93	30.4
9	29.16	38	27	43.58	-14.42	-49.5
10	118.26	38	93	107.75	10.51	8.9
11	39.62	60	27	49.29	-9.67	-24.4
12	45.68	71	27	51.57	-5.89	-12.9
13	149.34	71	100	134.48	14.86	10.0
14	41.97	82	5	15.60	26.37	62.8
15	59.48	93	27	55.47	4.01	6.7
16	148.58	93	93	137.15	11.43	7.7
17	163.14	93	100	144.64	18.51	11.3
18	73.14	100	38	72.64	0.50	0.7
19	114.06	100	71	114.79	-0.73	-0.6
20	153.44	100	93	139.86	13.58	8.9
						24.3 av

downwards. Did we make a mistake in arithmetic? No, there was no mistake, except in the selection of the power function in the first place. Figure 18 (the general shape of the power function for values of x greater than or equal to 0) could have told us that we would get what we did. This is another illustration of the value of the scatter diagram and a knowledge of the basic properties of the functional forms with which we are dealing. Consider a situation similar to this one except that the fit is better. In such a case we might well have used this relationship for extrapolating beyond the upper range of the sample.

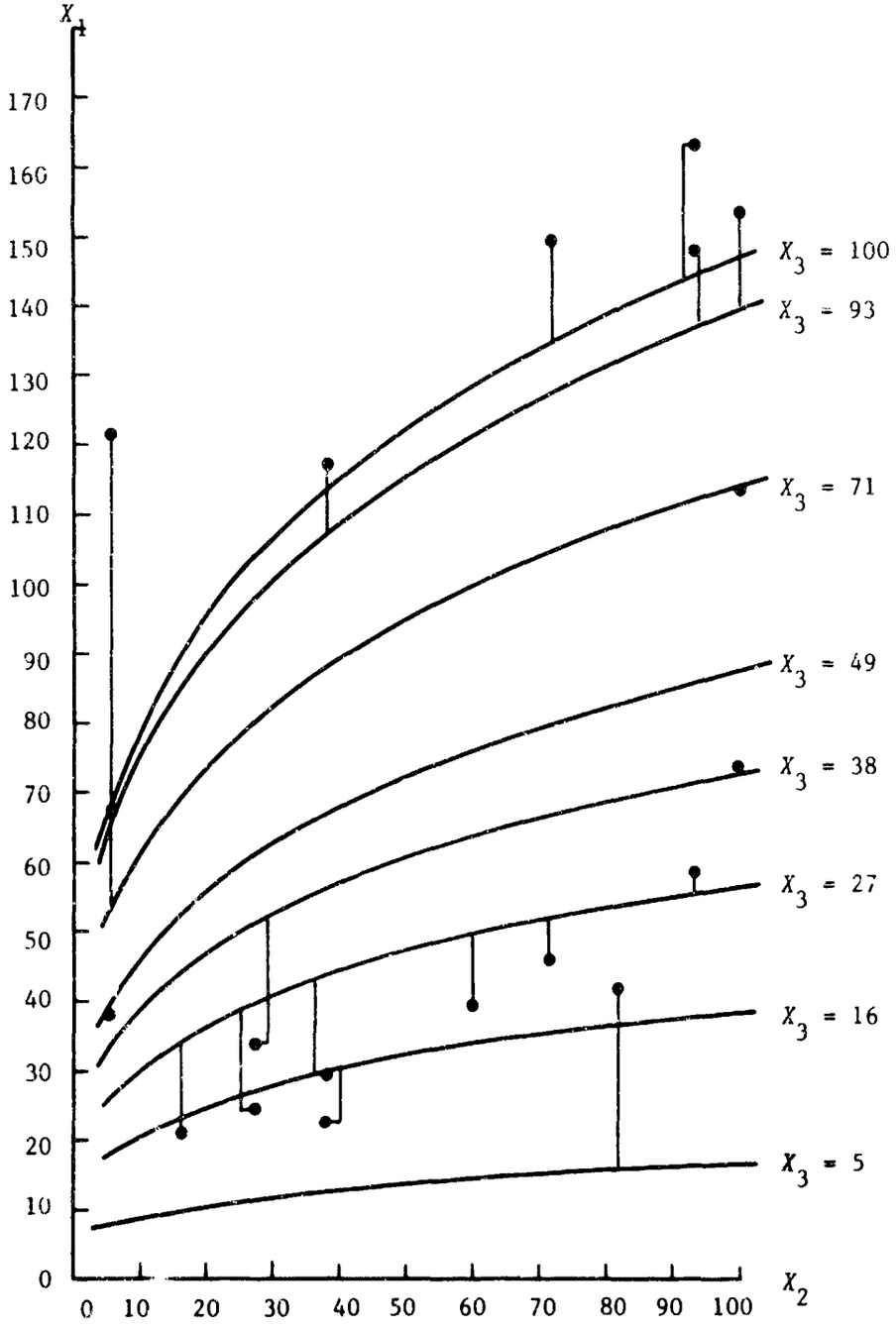


Fig. 44--Results of fitting a three-variable power function relationship ($\log X_1 = 0.16555 + 0.26963 \log X_2 + 0.73198 \log X_3$)

It is true, however, that the exponential has the general property we desire; refer back to Fig. 14. The two variable exponentials would be

$$X_1 = a_2^{X_2}$$

and

$$X_1 = a_3^{X_3},$$

or, in more useful form

$$X_1 = e^{a+b_2X_2},$$

and

$$X_1 = e^{a+b_3X_3}.$$

For further clarification on this point refer to the earlier section on the properties of the exponential.

When the natural logarithms of each side of each equation are taken, we have

$$\ln X_1 = a + b_2X_2$$

and

$$\ln X_1 = a + b_3X_3,$$

which combines into the following three-variable equation as before:

$$\ln X_1 = a + b_2X_2 + b_3X_3,$$

which is linear when the logarithm of X_1 is used in place of X_1.

The least-squares normal equations are as for the linear curve with $\ln X_1$ substituted for X_1:

$$\sum \ln X_1 = Na + b_2 \sum X_2 + b_3 \sum X_3,$$

$$\sum X_2 \ln X_1 = a \sum X_2 + b_2 \sum X_2^2 + b_3 \sum X_2 X_3,$$

$$\sum X_3 \ln X_1 = a \sum X_3 + b_2 \sum X_2 X_3 + b_3 \sum X_3^2.$$

The resulting equation is

$$\ln X_1 = 2.509 + 0.0092732\ X_2 + 0.019415\ X_3,$$

or

$$X_1 = e^{2.509 + 0.0092732\ X_2 + 0.019415\ X_3}.$$

Just how well this equation fits the data is shown in Table 17 and Fig. 45. We note from observing the scatter diagrams and the average percent deviations that the exponential relationship comes closer to fitting the data than does either the linear or the power function. The direction of curvature is as we predicted. However, while things are progressing, the exponential leaves much variation to be explained.

Another curve which, in general, has the desired properties (at least in part) is the parabola of the form

$$y = a + bx + cx^2.$$

The earlier section on the parabola provided a complete discussion of this equation. This equation is in two variables, however, and for our purposes we need one in three. Fortunately, we may proceed

Table 17

RESULTS OF FITTING A THREE-VARIABLE EXPONENTIAL RELATIONSHIP
($\ln X_1 = 2.509 + 0.0092732X_2 + 0.019415X_3$)

Observation	X_1	X_2	X_3	X_1(calc.)	$X_1 - X_1$(calc.)	Percent Deviation
1	7.31	5	5	14.19	-6.88	-94.1
2	37.67	5	49	33.34	4.33	11.5
3	67.37	5	71	51.10	16.27	24.2
4	121.31	5	100	89.74	31.57	26.0
5	20.93	16	27	24.08	-3.15	-15.1
6	24.77	27	27	26.67	-1.90	-7.7
7	33.57	27	38	33.02	0.55	1.6
8	22.78	38	16	23.86	-1.08	-4.7
9	29.16	38	27	29.54	-0.38	-1.3
10	118.26	38	93	106.38	11.88	10.1
11	39.62	60	27	36.22	3.40	8.6
12	45.68	71	27	40.11	5.57	12.2
13	149.34	71	100	165.49	-16.15	-10.8
14	41.97	82	5	28.98	12.99	31.0
15	59.48	93	27	49.19	10.29	17.3
16	148.58	93	93	177.15	-28.57	-19.2
17	163.14	93	100	202.94	-39.80	-24.4
18	73.14	100	38	64.98	8.16	11.2
19	114.06	100	71	123.32	-9.26	-8.1
20	153.44	100	93	189.03	-35.59	-23.2
						18.1 av

as before. The two variable equations are:

$$X_1 = a + b_2X_2 + c_2X_2^2,$$

and

$$X_1 = a + b_3X_3 + c_3X_3^2,$$

which combined, form

$$X_1 = a + b_2X_2 + c_2X_2^2 + b_3X_3 + c_3X_3^2.$$

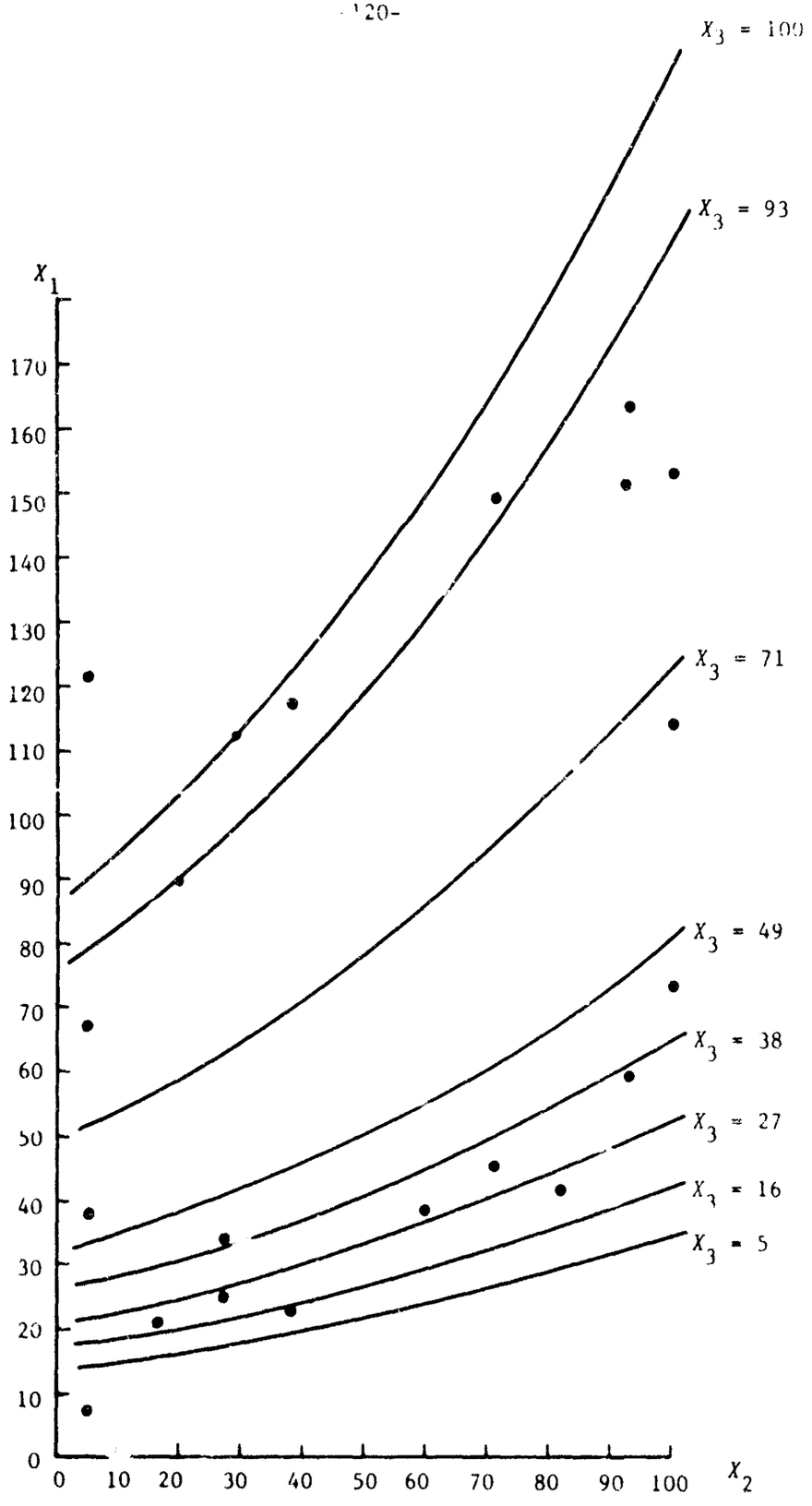


Fig. 45--Results of fitting a three-variable exponential relationship ($\ln X_1 = 2.509 + 0.0092732X_2 + 0.019415X_3$)

Notice that instead of two independent variables, X_2 and X_3, we now have four variables: X_2, X_2^2, X_3, and X_3^2. Fortunately X_2^2 and X_3^2 may be calculated given X_2 and X_3, so that we have a special case of fitting what is essentially a five-variable linear relationship. The least-squares normal equations follow:

$$\sum X_1 = Na + b_2 \sum X_2 + c_2 \sum X_2^2 + b_3 \sum X_3 + c_3 \sum X_3^2,$$

$$\sum X_1X_2 = a \sum X_2 + b_2 \sum X_2^2 + c_2 \sum X_2^3 + b_3 \sum X_2X_3 + c_3 \sum X_3^2,$$

$$\sum X_1X_2^2 = a \sum X_2^2 + b_2 \sum X_2^3 + c_2 \sum X_2^4 + b_3 \sum X_2^2X_3 + c_3 \sum X_2^2X_3^2,$$

$$\sum X_1X_3 = a \sum X_3 + b_2 \sum X_2X_3 + c_2 \sum X_2^2X_3 + b_3 \sum X_3^2 + c_3 \sum X_3^3,$$

$$\sum X_1X_3^2 = a \sum X_3^2 + b_2 \sum X_2X_3^2 + c_2 \sum X_2^2X_3^2 + b_3 \sum X_3^3 + c_3 \sum X_3^4.$$

Manual solution of this set of equations is lengthy at best. Consequently, one of the many computer programs available should probably be used. With a computer, the task becomes a simple one, and the chance of making errors in arithmetic is minimum. In the case of our example the derived equation is

$$X_1 = 5.006 + 0.2498\ X_2 + 0.002301\ X_2^2 + 0.1499\ X_3 + 0.01000\ X_3^2.$$

Table 18 and Fig. 46 indicate that we have indeed found the correct empirical equation. However, even with fits as good as this one, unless there is a logical base for the particular equation, extrapolations beyond the range of the data should be made with extreme caution. Such is particularly true when the relationship is a parabola. (Review the section of this Memorandum on the general properties of parabolas.)

Table 18

RESULTS OF FITTING A THREE-VARIABLE PARABOLIC RELATIONSHIP
($X_1 = 5.006 + 0.2498X_2 + 0.002301X_2^2 + 0.1499X_3 + 0.0100X_3^2$)

Observation	X_1	X_2	X_3	X_1(calc.)	X_1-X_1(calc.)	Percent Deviation
1	7.31	5	5	7.31	---	---
2	37.67	5	49	37.67	---	---
3	67.37	5	71	67.37	---	---
4	121.31	5	100	121.31	---	---
5	20.93	16	27	20.93	---	---
6	24.77	27	27	24.77	---	---
7	33.57	27	38	33.57	---	---
8	22.78	38	16	22.78	---	---
9	29.16	38	27	29.16	---	---
10	118.26	38	93	118.26	---	---
11	39.62	60	27	39.62	---	---
12	45.68	71	27	45.68	---	---
13	149.34	71	100	149.34	---	---
14	41.97	82	5	41.97	---	---
15	59.48	93	27	59.48	---	---
16	148.58	93	93	148.58	---	---
17	163.14	93	100	163.14	---	---
18	73.14	100	38	73.14	---	---
19	114.06	100	71	114.06	---	---
20	153.44	100	93	153.44	---	---

We have illustrated the combination of two similar two-variable relationships to form a single three-variable relationship. In fact, certain dissimilar two-variable relationships may also be combined. For example,

$$X_1 = a + b_2X_2$$

may be combined with

$$X_1 = a + b_3X_3 + c_3X_3^2$$

to form

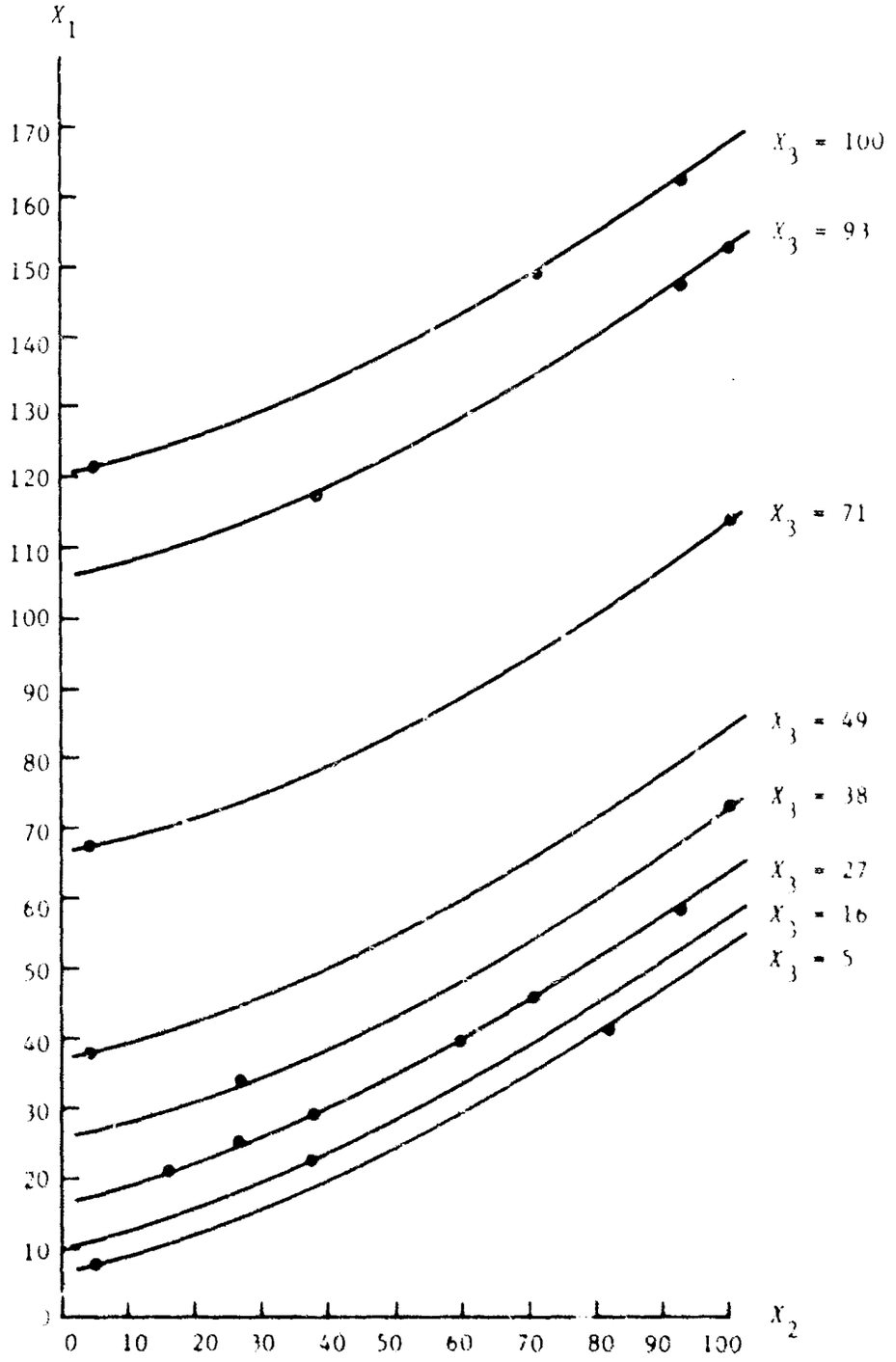


Fig. 46--Results of fitting a three-variable parabolic relationship ($X_1 = 5.006 + 0.2498X_2 + 0.002301X_2^2 + 0.1499X_3 + 0.0100X_3^2$)

$$X_1 = a + b_2X_2 + b_3X_3 + b_3X_3^2.$$

If it were observed that X_1 varied linearly with X_2 and nonlinearly with X_3, an equation like the one above might then be an appropriate choice.

Appendix A

DERIVATION OF THE NORMAL EQUATIONS FOR A LEAST-SQUARES FIT OF A STRAIGHT LINE, A PARABOLA, AND A THREE-VARIABLE LINEAR EQUATION

A Straight Line

In curve fitting the general equation of a straight line is

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the parameters to be determined such that the sum of the squares of the deviations from the resulting line is a minimum. The carets are placed over those values that are to be estimates. If we let each value of the dependent y variable be represented by y_i with the subscript assigned according to the data point we are using, we can write the general formula for the deviations d as

$$d_i = y_i - \hat{y}_i \ .$$

On substituting the expression for $\hat{y}$ we have

$$d_i = y_i - (\hat{\alpha} + \hat{\beta}x_i) \ .$$

The squared deviations are

$$d_i^2 = [y_i - (\hat{\alpha} + \hat{\beta}x_i)]^2 \ ,$$

which on expansion becomes

$$d_i^2 = y_i^2 - 2\hat{\alpha}y_i - 2\hat{\beta}x_iy_i + \hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta}x_i + \hat{\beta}^2x_i \ .$$

We need such an expression because our interest is in minimizing

the sum of the squared deviations. The expression that follows represents symbolically the summation of the above expression across all values of i from 1 to n, where 1 would represent the first data point, 2 the second, and n the last:

$$Q = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} y_i^2 - 2\hat{\alpha}\sum_{i=1}^{n} y_i - 2\hat{\beta}\sum_{i=1}^{n} x_i y_i + \hat{\alpha}^2 n + 2\alpha\hat{\beta}\sum_{i=1}^{n} x_i + \hat{\beta}^2 \sum_{i=1}^{n} x_i .$$

From calculus we know that, for this expression to be a minimum, the partial derivatives of Q taken with respect to $\hat{\alpha}$ and $\hat{\beta}$ must be equal to 0. It can also be shown that this is a *sufficient* condition for the above expression to be a minimum.* Thus the procedure is to obtain these two partial derivatives and to equate them to 0. The partial derivatives are

$$\frac{\partial\left(\Sigma d_i^2\right)}{\partial\alpha} = -2\sum_{i=1}^{n} y_i + 2n\hat{\alpha} + 2\hat{\beta}\sum_{i=1}^{n} x_i ;$$

$$\frac{\partial\left(\Sigma d_i^2\right)}{\partial\beta} = -2\sum_{i=1}^{n} x_i y_i + 2\hat{\alpha}\sum_{i=1}^{n} x_i + 2\hat{\beta}\sum_{i=1}^{n} x_i^2 .$$

After equating each of these to 0 and simplifying, we have

$$\sum_{i=1}^{n} y_i = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} x_i ;$$

$$\sum_{i=1}^{n} x_i y_i = \hat{\alpha}\sum_{i=1}^{n} x_i + \hat{\beta}\sum_{i=1}^{n} x_i^2 ,$$

which are the required normal equations. All of the information indicated both by the summation signs and by n can be determined directly

*The condition of *sufficiency* applies to any function that is linear with respect to all of its parameters, such as the parabola.

from the data; this will result in two equations in two unknowns (α and γ) which can be solved for simultaneously.

A Parabola

The general equation of a parabola is

$$y = \hat{\alpha} + \hat{\beta}x + \hat{\gamma}x^2,$$

where $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are the parameters to be determined.

An individual squared deviation may be represented as follows:

$$d_i^2 = [y_i - (\hat{\alpha} + \hat{\beta}x_i + \hat{\gamma}x_i^2)]^2,$$

which when expanded is

$$d_i^2 = y_i^2 - 2\hat{\alpha}y_i - 2\hat{\beta}x_iy_i - 2\hat{\gamma}x_i^2y_i + \hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta}x_i + 2\hat{\alpha}\hat{\gamma}x_i^2 + \hat{\beta}^2x_i^2 + 2\hat{\beta}\hat{\gamma}x_i^3 + \hat{\gamma}^2x_i^4.$$

The sum of the squared deviations taken from $i = 1$ to n is

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} y_i^2 - 2\hat{\alpha}\sum_{i=1}^{n} y_i - 2\hat{\beta}\sum_{i=1}^{n} x_iy_i - 2\hat{\gamma}\sum_{i=1}^{n} x_i^2y_i + n\hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta}\sum_{i=1}^{n} x_i + 2\hat{\alpha}\hat{\gamma}\sum_{i=1}^{n} x_i^2 + \hat{\beta}^2\sum_{i=1}^{n} x_i^2 + 2\hat{\beta}\hat{\gamma}\sum_{i=1}^{n} x_i^3 + \hat{\gamma}^2\sum_{i=1}^{n} x_i^4.$$

To minimize, we take the partial derivatives with respect to $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ and equate them to 0 as follows:

$$\frac{\partial d^2}{\partial \hat{\alpha}} = -2\sum_{i=1}^{n} y_i + 2n\hat{\alpha} + 2\hat{\beta}\sum_{i=1}^{n} x_i + 2\hat{\gamma}\sum_{i=1}^{n} x_i^2,$$

$$\frac{\partial d^2}{\partial \hat{\beta}} = -2\sum_{i=1}^{n} x_i y_i + 2\hat{\alpha}\sum_{i=1}^{n} x_i + 2\hat{\beta}\sum_{i=1}^{n} x_i^2 + 2\hat{\gamma}\sum_{i=1}^{n} x_i^3,$$

$$\frac{\partial d^2}{\partial \gamma} = -2\sum_{i=1}^{n} x_i^2 y_i + 2\hat{\alpha}\sum_{i=1}^{n} x_i^2 + 2\hat{\beta}\sum_{i=1}^{n} x_i^3 + 2\hat{\gamma}\sum_{i=1}^{n} x_i^4,$$

$$\sum_{i=1}^{n} y_i = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} x_i + \hat{\gamma}\sum_{i=1}^{n} x_i^2,$$

$$\sum_{i=1}^{n} x_i y_i = \hat{\alpha}\sum_{i=1}^{n} x_i + \hat{\beta}\sum_{i=1}^{n} x_i^2 + \hat{\gamma}\sum_{i=1}^{n} x_i^3,$$

$$\sum_{i=1}^{n} x_i^2 y_i = \hat{\alpha}\sum_{i=1}^{n} x_i^2 + \hat{\beta}\sum_{i=1}^{n} x_i^3 + \hat{\gamma}\sum_{i=1}^{n} x_i^4.$$

These are the normal equations for fitting a parabola using the least-squares criterion. The sums and sums of products are calculated directly from the data and substituted into the normal equations leaving three equations and three unknowns. These three equations are solved simultaneously for $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$. As for the straight line, the solutions are unique and exact for all x and y.

A Three-variable Linear Equation

The general form of the linear three-variable equation is

$$\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

The derivation procedure is identical to that used in deriving the normal equations for the straight line and for the parabola. The squared deviations are

$$d^2 = [y - (\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2)]^2.$$

The above expression is expanded and summed across all the data points, and the partial derivatives of the summation equation with respect to α, β_1, and β_2 are taken and equated to 0. The resulting normal equations are

$$\sum y = n\hat{\alpha} + \hat{\beta}_1 \sum x_1 + \hat{\beta}_2 \sum x_2,$$

$$\sum x_1 y = \hat{\alpha} \sum x_1 + \hat{\beta}_1 \sum x_1^2 + \hat{\beta}_2 \sum x_1 x_2,$$

$$\sum x_2 y = \hat{\alpha} \sum x_2 + \hat{\beta}_1 \sum x_1 x_2 + \hat{\beta}_2 \sum x_2^2.$$

In the above equation the subscripts are used to distinguish between the two independent variables and their coefficients rather than to indicate the range of summation as before. Although it is not specifically indicated here, it should be understood that the sums are to be taken across all data points.

Appendix B

DERIVATION OF THE FORMULA FOR CALCULATING THE CONSTANT x

The value of x, as is shown in Fig. B-1, must be such that when the values of the y coordinates of points on L_1 are reduced by that amount, the new points fall on the line L_2. Further, L_2 must be linear in terms of logarithms.

For L_2 to be linear in terms of logarithms, the triangles ABC and BDE must be similar. In other words, L_2 must have a constant slope. It is this fact that provides the basis for calculating x.

The slope of the triangle ABC is equal to

$$\frac{CA}{BC},$$

and the slope of the triangle BDE is equal to

$$\frac{EB}{DE}.$$

Also the two slopes must be equal to each other, e.g.,

$$\frac{CA}{BC} = \frac{EB}{DE}. \tag{1}$$

When the coordinates of the appropriate points are used to calculate the lengths of the above line segments and the results are substituted in Eq. 1, we have

$$\frac{\log(y_1 - x) - \log(y_3 - x)}{\log x_1 - \log x_3} = \frac{\log(y_3 - x) - \log(y_2 - x)}{\log x_3 - \log x_2}. \tag{2}$$

Since we are free to select the three points (x_1, y_1), (x_2, y_2) and (x_3, y_3) in any way we wish, we do so in such a way that the denominators

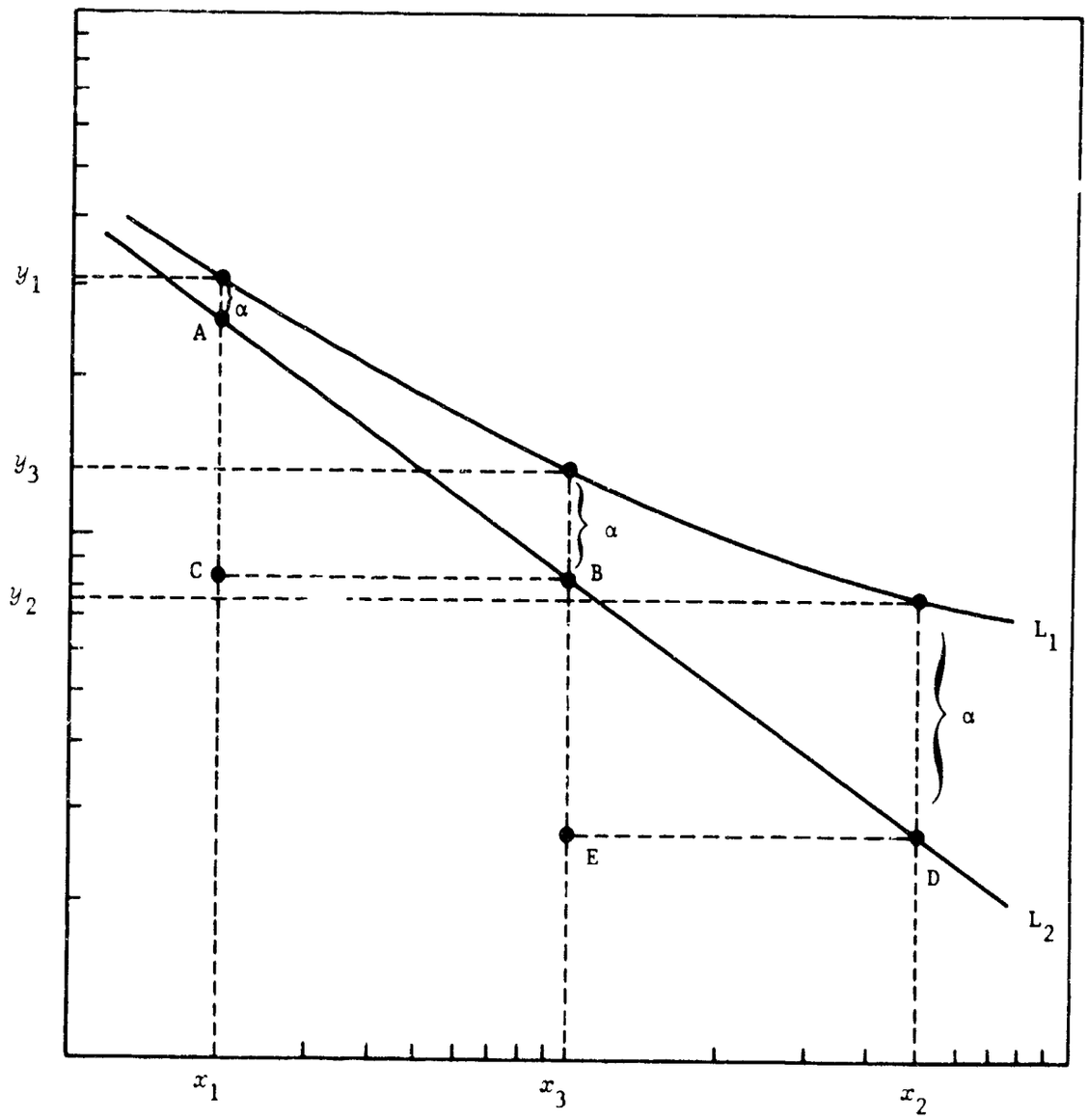


Fig. B-1--Determining the value of the constant α

of the two fractions in Eq. 2 are equal, such as

$$\log x_1 - \log x_3 = \log x_3 - \log x_2. \tag{3}$$

General practice is to choose x_1 and x_2 at the extremities of L_1 and to let Eq. 3 determine the value of x_3, such as

$$2 \log x_3 = \log x_1 + \log x_2,$$

or

$$\log x_3 = \frac{\log x_1 + \log x_2}{2}. \tag{4}$$

As can be seen, $\log x_3$ is the average of, or half way between, $\log x_1$ and $\log x_2$.

Equation 4 in arithmetic form is

$$x_3 = \sqrt{x_1 x_2};$$

x_3 is seen to be the geometric mean of x_1 and x_2.

If x_3 is chosen in this way, Eq. 2 then reduces to

$$\log(y_1 - a) - \log(y_3 - a) = \log(y_3 - a) - \log(y_2 - a).$$

In arithmetic form we have

$$\frac{y_1 - a}{y_3 - a} = \frac{y_3 - a}{y_2 - a};$$

$$(y_1 - a)(y_2 - a) = (y_3 - a)^2;$$

$$y_1 y_2 - a y_2 - a y_1 + a^2 = y_3^2 - 2 a y_3 + a^2;$$

$$\alpha = \frac{y_1 y_2 - y_3^2}{y_1 + y_2 - 2y_3}. \qquad (5)$$

Equation 5 is the desired result.

If we had been concerned with the semi-log case, Fq. 2 would have been

$$\frac{\log(y_1 - \alpha) - \log(y_3 - \alpha)}{x_1 - x_3} = \frac{\log(y_3 - \alpha) - \log(y_2 - \alpha)}{x_3 - x_2};$$

and Eq. 4 would be

$$x_3 = \frac{x_1 + x_2}{2}.$$

We would therefore make x_3 the average of x_1 and x_2 instead of the geometric mean. Equation 5 applies as before.

BIBLIOGRAPHY

A Selected List of Recommended Readings

Boren, H. E., *Curves: A Five-function Curve-fitting Computer Program*, The RAND Corporation, RM-5762-PR, November, 1968.

Croxton, F. E. and D. J. Cowden, *Applied General Statistics*, 3rd ed., Prentice-Hall, Inc., New York, 1960.

Davis, D. S., *Nomography and Empirical Equations*, Reinhold, New York, 1955.

Ezekiel, M. and K. A. Fox, *Methods of Correlation and Regression Analysis*, Wiley and Sons, New York, 1959.

Fowler, F. P. and E. W. Sandberg, *Basic Mathematics for Administration*, Wiley and Sons, New York, 1962.

Graver, C. A. and H. E. Boren, *Multivariate Logarithmic and Exponential Regression Models*, The RAND Corporation, RM-4879-PR, July 1967.

Levens, A. S., *Nomography*, Wiley and Sons, New York, 1959.

Middlemiss, R. R., *Analytic Geometry*, McGraw-Hill, New York, 1955.

Spurr, W. A. and C. P. Bonini, *Statistical Analysis for Business Decisions*, Richard D. Irwin, Inc., Homewood, Illinois, 1967.

DOCUMENT CONTROL DATA

1 ORIGINATING ACTIVITY	2a REPORT SECURITY CLASSIFICATION
THE RAND CORPORATION	UNCLASSIFIED
	2b. GROUP

3 REPORT TITLE
SOME CURVE-FITTING FUNDAMENTALS

4. AUTHOR(S) (Last name, first name, initial)

Petruschell, R. L.

5. REPORT DATE	6a. TOTAL No. OF PAGES	6b. No. OF REFS.
December 1968	147	---

7. CONTRACT OR GRANT No.	8. ORIGINATOR'S REPORT No.
DAHC15 67 C 0150	RM-5766-SA

9a. AVAILABILITY / LIMITATION NOTICES	9b. SPONSORING AGENCY
DDC-1	Office of the Assistant Secretary of Defense (Systems Analysis)

10. ABSTRACT

A description of the curve-fitting process for the cost analyst. The study is characterized by intuitive discussions with illustrations of computational procedures, and treats the more complex relationships of cost analysis by an approach that integrates analytic geometry with curve-fitting methods. In order to develop an equation to describe a particular relationship, the approach combines the properties of specific functional forms--the straight line, the exponential, the power function, and the parabola--with the values of equation constants. Examples of curves fit to two-variable and multi-variable relationships are shown. Both linear and nonlinear cases are included.

11. KEY WORDS

Cost analysis
Cost estimating relationships
Curve fitting
Statistical methods and processes