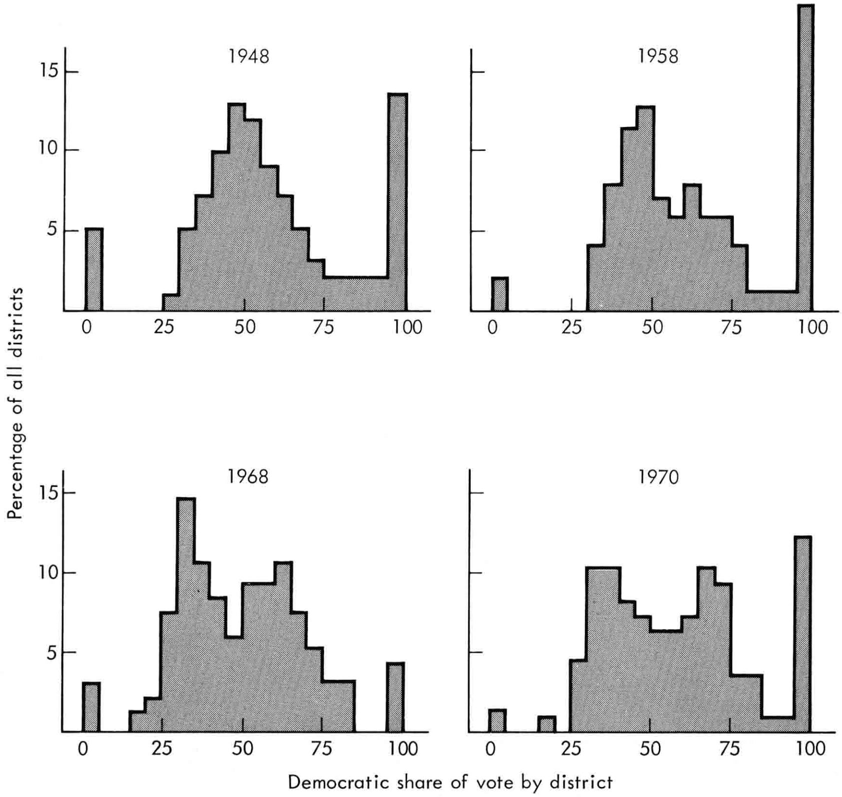


FIGURE 3-16 Distribution of congressional vote by district

contested race in all 24 congressional districts in 1970 was a 54–46 division of the vote; in contrast, in 1960, seven districts had closer races than that. The closest 1970 race in Pennsylvania was 55–45; in Ohio, 53–47.

In conclusion, then, we have seen here how the linear regression model can be used to measure two important qualities of an electoral system—the responsiveness and the partisan bias of the system. These two measurements might even be used by the courts to evaluate the fairness and the effectiveness of redistricting plans submitted to the courts.

This example has shown the economy of the regression model, in which the estimate of the slope takes us quickly to the central political issues in the data. There was little to learn from a correlation coefficient in this case (and in many others), for we already knew that there was a strong relationship between how many votes and how many seats a party received. In contrast to the correlation coefficient, the regression model gave us a measure permitting politically meaningful comparisons across different political systems. Note also that a correlational analysis misses the method of assessing the partisan bias—an estimate which flows naturally from the regression model. Finally, look back at those four histograms in Figure 3-16. Note how informative they are with respect to the performance of the electoral system and how directly they make the point. Such is generally the case. Pictures of the data—charts, scatterplots, histograms, or just the values of a variable marked out on a line—are powerful aids to analysis. They also are easy to produce, either by hand or by computer.

Example 5: Comparing the Slope and the Correlation Coefficient

Both the correlation coefficient, *r*, and the slope of the fitted line, $\hat{\beta}_1$, are numerical summaries of the relationship between two variables. The slope, since it expresses the relationship in terms of the units in which X and Y are measured, is often a more useful summary measure than the correlation. This was true in the examples dealing with midterm congressional elections and the translation of votes into seats. In those examples the slope carried the important message in the data. Such interpretations of the slope require, however, that the units of measurement of the X and Y variables make some sort of interpretative sense.

For example, in examining responses to an interview questionnaire—and correlating relationships over the different responses to questions—it is difficult to interpret a measure of the rate of change on the intensity of feeling on one question with respect to the intensity of feeling on another. In such a case, the correlation coefficient may be more appropriate.

John Tukey has expressed these views strongly:

> . . . [M]ost correlation coefficients should never be calculated. . . . [C]orrelation coefficients are justified in two and only two circumstances, when they are regression coefficients, or when measurement of one or both variables on a determinate scale is hopeless. . . . The other area in which correlation coefficients are prominent

> includes psychometrics and educational testing in general. This is surely a situation where determinate scales are hopeless.[11]

The correlation coefficient, r, can be interpreted in a number of ways. Its square, r^2, is the proportion of variance in the response variable explained by the describing variable. Or it can be viewed as the average covariation of standardized variables:

$$r = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right).$$

That is, each observation is rescaled and measured in terms of how many standard deviations it is from the mean—for a given observation (X_i, Y_i):

$$\frac{X_i - \bar{X}}{s_X} \quad \text{and} \quad \frac{Y_i - \bar{Y}}{s_Y}.$$

The product of the rescaled variables is averaged over all observations to yield the correlation coefficient.

Both the correlation coefficient and the slope can be dominated by a few extreme values in the data. Since we are working with products of deviations from the mean, a data point far from the mean on both variables can virtually determine the value of r and β_1. Thus sometimes r and β_1 do not provide very good summaries of the relationship between X and Y. They fail when the relationship is nonlinear and when the data contain extreme outlying values.[12] The problems are easily detected from a scatterplot of the data. Thus one practical moral is that every calculation of r and β_1 should also involve an inspection of the scatterplot.

Let us now look at a series of scatterplots. First are examples in which the data are well described by the linear model: the data are

[11] J. W. Tukey, "Causation, Regression, and Path Analysis," in O. Kempthorne, et al., eds., *Statistics and Mathematics in Biology* (Ames, Iowa: Iowa State College Press, 1956), pp. 38–39.

[12] In the case of many nonlinear scatterplots, the data can be transformed and the linear model estimated. Outliers can be treated by transformations, by removing them from the analysis, or by "Winsorizing" them (setting the most extreme value on a variable to the next most extreme). See Joseph B. Kruskal, "Special Problems of Statistical Analysis: Transformations of Data," *International Encyclopedia of the Social Sciences* (New York: Macmillan, 1968), vol. 15, 182–93; and F. J. Anscombe, "Outliers," *ibid.*, 178–82.

roughly oriented around a straight line with no extreme outliers (Figure 3-17).

We finally turn to some data sets for which the correlation and the fitted line fail to summarize the data effectively. Figure 3-18 shows three scatterplots with widely divergent patterns of relationship between X and Y. The first plot shows no relationship, discounting the one extreme outlier on both measures. The second plot suggests a moderately strong linear relationship between X and Y. The third plot reveals a rather marked curvilinear relationship between X and Y, revealing that as X increases, Y gets bigger even faster. Despite the great variation in the visual message, *the correlation between X and Y* is the same in all three cases. Also, the slopes do not differ greatly in the three cases.

Often a set of data for which the linear model is not immediately applicable can be transformed so the linear model is valuable. Or, to put it the other way around: many models with nonlinearities in the variables can be estimated by so-called "linear" regression.

For example, suppose we work with the logarithm of the one of the variables and have the model

$$Y = \beta_0 + \beta_1 \log X.$$

This model is estimated by letting $X' = \log X$ and then performing the usual least-squares regression for the model

$$Y = \beta_0 + \beta_1 X'.$$

Thus the criticism sometimes made that linear regression "assumes linearity" is a bit misleading, since the assumption can, in fact, be checked—and, if false, the model then redesigned for purposes of estimation. In fact, a better name for what this chapter has been all about is "fitting curves to relationships between two variables."

In summary, then, fitting lines to relationships between variables is often a useful and powerful method of summarizing a set of data. Regression analysis fits naturally with the development of causal explanations, simply because the research worker must, at a minimum, know what he or she is seeking to explain. The regression model is surprisingly flexible; and we now illustrate methods that increase its range of application.

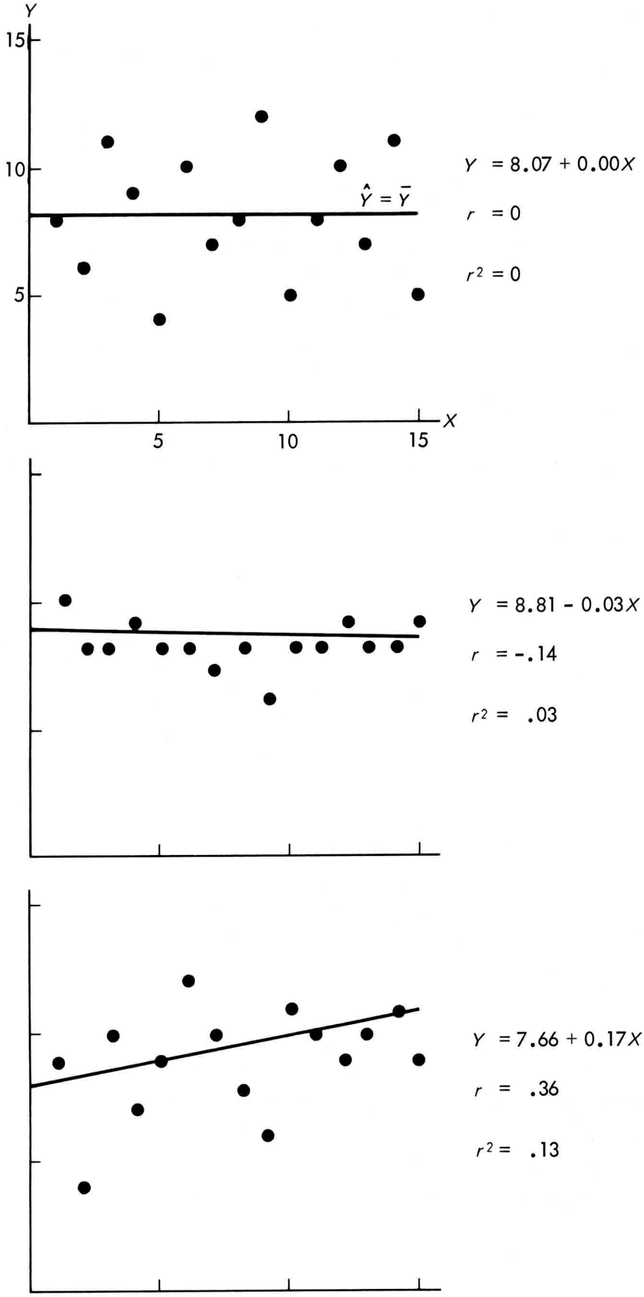


FIGURE **3-17** Data relatively well described by a fitted line

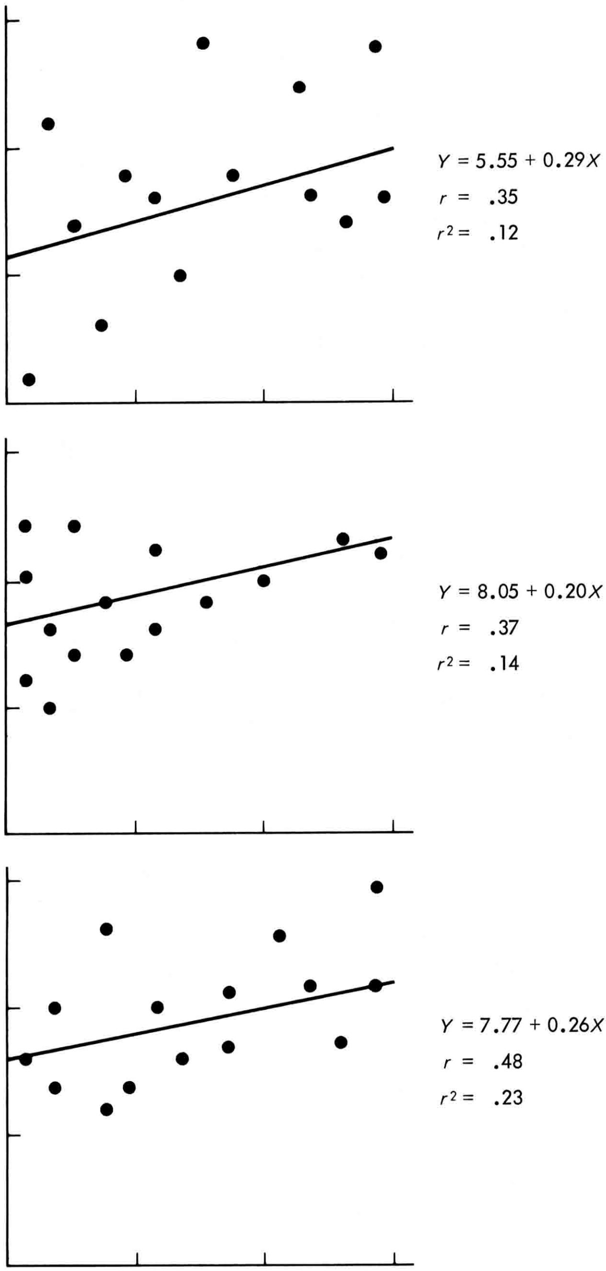


FIGURE 3–17 Data relatively well described by a fitted line

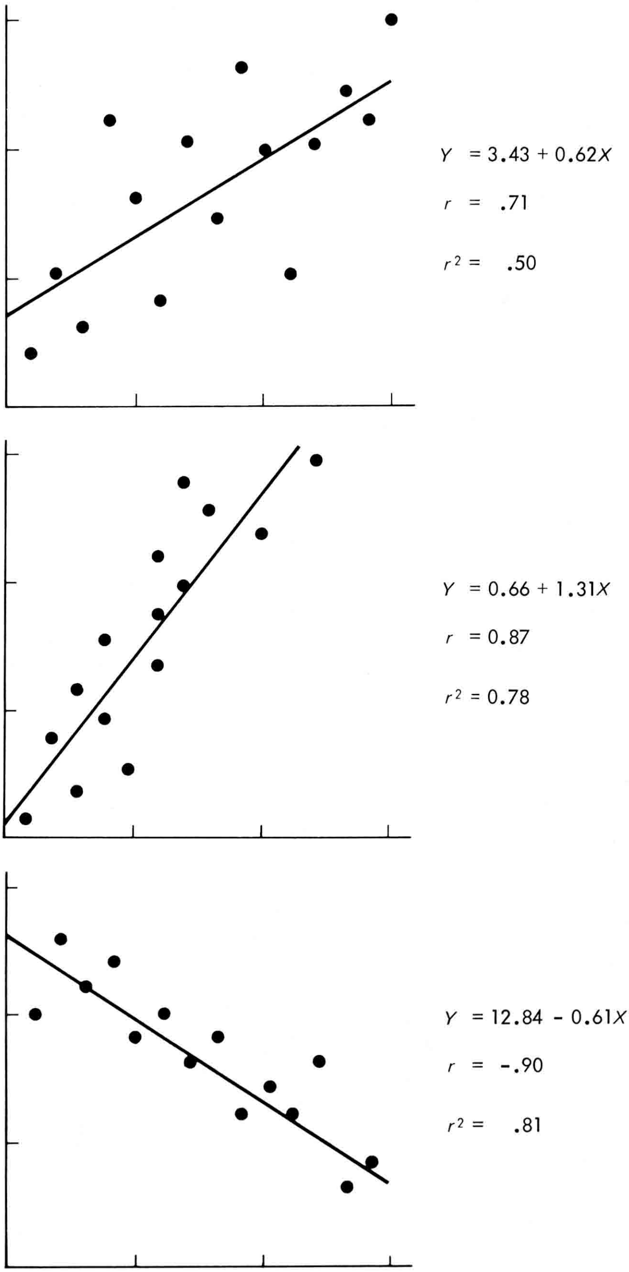


FIGURE 3-17 Data relatively well described by a fitted line

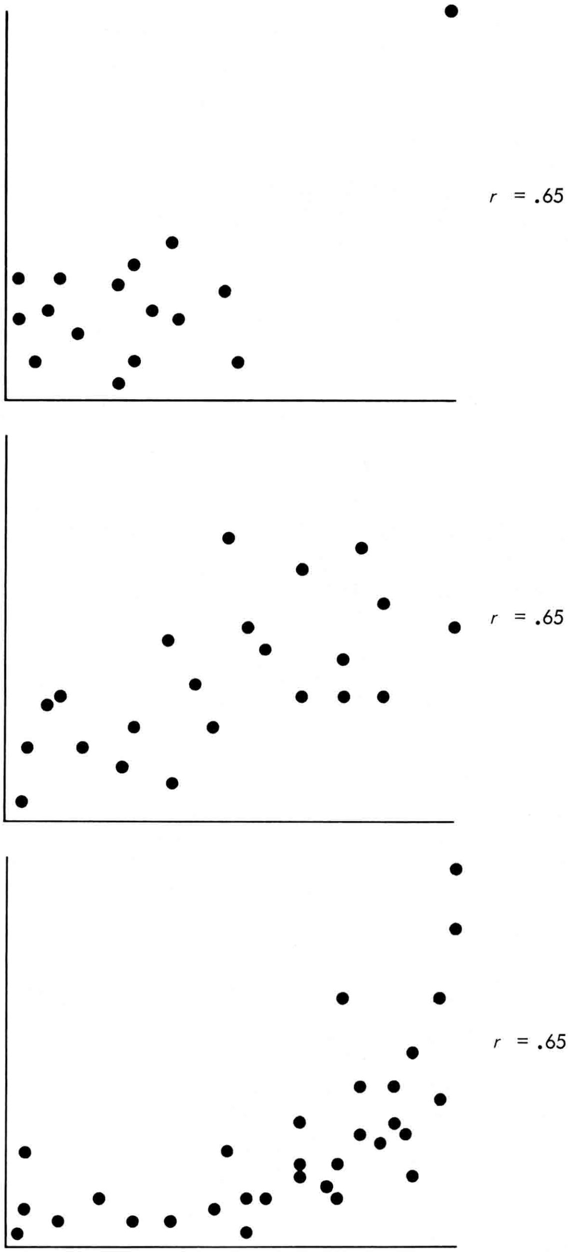


FIGURE 3-18 Three scatterplots with the same correlation

Example 6: Interpretation of Regression Coefficients when the Variables are Re-expressed as Logarithms (with Five Examples)

> Data that are *counts* of populations, vital statistics, census data, and the like are almost always improved by taking logs. . . . Charles Winsor frequently prescribed the taking of logs of all naturally occurring counts (plus *one,* to handle that embarrassing quantity *zero*) before analyzing them—no matter what the sources [of the data].[13]

Often the logarithm of a variable is taken before entering that variable in a regression analysis. The logarithmic transformation serves several purposes:

1. The resulting regression coefficients sometimes have a more useful theoretical interpretation compared to a regression based on unlogged variables.
2. Badly skewed distributions—in which many of the observations are clustered together combined with a few outlying values on the scale of measurement—are transformed by taking the logarithm of the measurements so that the clustered values are spread out and the large values pulled in more toward the middle of the distribution.
3. Some of the assumptions underlying the regression model and the associated significance tests are better met when the logarithm of the measured variables is taken.

REMEMBERING LOGARITHMS

The logarithm to the base b of a number x, written as $\log_b x$, is the power to which the base must be raised to yield x. Thus

$\log_{10} 1000 = 3$, because $10^3 = 1000$.

Similarly:

$\log_{10} 10{,}000 = 4$,	because $10^4 = 10{,}000$.
$\log_{10} 1 = 0$,	because $10^0 = 1$.
$\log_{10} 2 = .30103$,	because $10^{.30103} = 2$.
$\log_{10} 2000 = 3.30103$,	because $10^{3.30103} = 200$.
$\log_{10} 20{,}000 = 4.30103$,	because $10^{4.30103} = 20{,}000$.

In short, then, logarithms are powers of the base. The base 10, the base e (which forms what are called "natural" logarithms), and

[13] Forman S. Acton, *Analysis of Straight-Line Data* (New York: Wiley, 1959), p. 223.

the base 2 are the ones most commonly used. Logs to the base 2 take the following form:

$\log_2 8 = 3$, because $2^3 = 8$.

The logarithm of zero does not exist (regardless of the base) and therefore must be avoided. In logging variables with some zero values (especially those deriving from counts), the most common procedure is to add one to all the observations of the variable.

Finally, we should recall the following rules for manipulation of logarithms:

For $x > 0$ and $y > 0$:

$$\log xy = \log x + \log y.$$

For example,

$$\begin{aligned}\log 20{,}000 &= \log (2)(10{,}000)\\ &= \log 2 + \log 10{,}000\\ &= .30103 + 4\\ &= 4.30103.\end{aligned}$$

$$\log \frac{x}{y} = \log x - \log y.$$

$$\log x^n = n \log x.$$

Let us first look at the effect of taking logarithms on the measurement scale of a single variable. Figure 3-19 shows the relationship between X and $\log X$; and Table 3-6 (page 111) tabulates the populations of some 29 countries of the world along with the logarithm of population. Note how the logarithmic transformation pulls the extremely large values in toward the middle of the scale and spreads the smaller values out in comparison to the original, unlogged values of the variable. Although the transformation preserves the rank ordering of the countries with respect to population, it still does produce quite a major change in the scaling of the variable here: the correlation between the population and the logarithm of population for the 29 countries is .68.

One reason for expressing population size here as a power of ten (that is, logging size to the base ten) is simply for convenience: if our scatterplots are going to include and differentiate between Iceland and Norway as well as the United States and India, then something must be done to compress the extreme end of the distribution. Logging

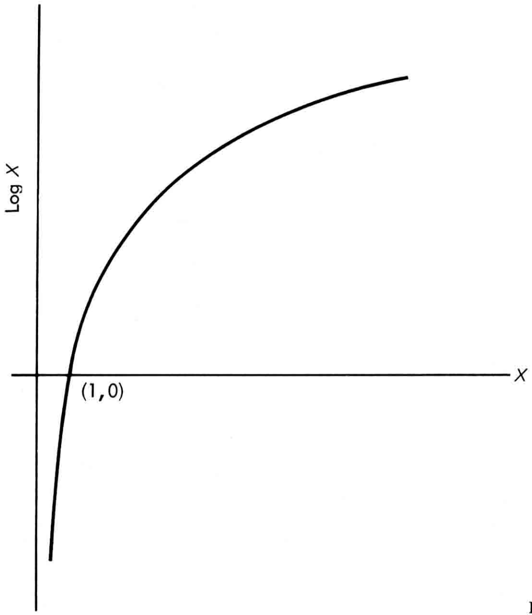


FIGURE 3-19 X vs. log X

size transforms the original skewed distribution into a more symmetrical one by pulling in the long right tail of the distribution toward the mean. The short left tail is, in addition, stretched. The shift toward symmetrical distribution produced by the log transform is not, of course, merely for convenience. Symmetrical distributions, especially those that resemble the normal distribution, fulfill statistical assumptions that form the basis of statistical significance testing in the regression model. Figure 3-20 shows the contrast between the logged and unlogged frequency distributions of population.

Logging skewed variables also helps to reveal the patterns in the data. Figure 3-21 shows the relationship between the population size of a country and the size of its parliament—for the unlogged and the logged variables. Note how the rescaling of the variables by taking logarithms reduces the nonlinearity in the relationship and removes much of the clutter resulting from the skewed distributions on both variables; in short, the transformation helps clarify the relationship between the two variables. It also, as we will see now, leads to a theoretically meaningful regression coefficient.

Much of the value of the logarithmic transformation derives from its contribution to the testing of theoretical models by means of linear

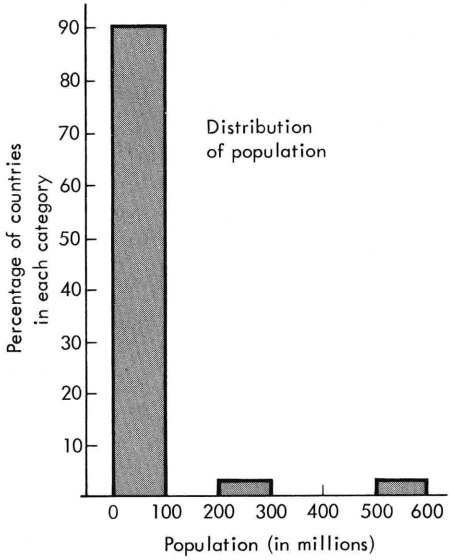
TABLE 3-6
Population, 29 Countries, 1970

Country	*Population*	*Log (Population)*
Iceland	200,000	5.30
Luxembourg	400,000	5.60
Trinidad and Tobago	1,100,000	6.04
Costa Rica	1,800,000	6.25
Jamaica	2,000,000	6.30
New Zealand	2,800,000	6.45
Lebanon	2,800,000	6.45
Israel	2,900,000	6.46
Uruguay	2,900,000	6.46
Ireland	3,000,000	6.48
Norway	3,900,000	6.59
Finland	4,700,000	6.67
Denmark	4,900,000	6.69
Switzerland	6,300,000	6.80
Austria	7,400,000	6.87
Sweden	8,000,000	6.90
Belgium	9,700,000	6.99
Chile	9,800,000	6.99
Australia	12,500,000	7.10
Netherlands	13,000,000	7.11
Canada	21,400,000	7.33
Philippines	38,100,000	7.58
France	51,100,000	7.71
Italy	53,700,000	7.73
United Kingdom	56,000,000	7.75
West Germany	58,500,000	7.77
Japan	103,500,000	8.02
United States	204,600,000	8.31
India	554,600,000	8.74

regression.[14] In interpreting regression coefficients of such models when the variables are logged, we have the following possibilities:

		Describing variable (X)	
		Logged	*Not logged*
Response variable (Y)	*Logged*	I	II
	Not logged	III	IV

[14]For further information see J. Johnston, *Econometric Methods,* 2d ed. (New York: McGraw-Hill, 1972), chap. 3; N. R. Draper and H. Smith, *Applied Regression*



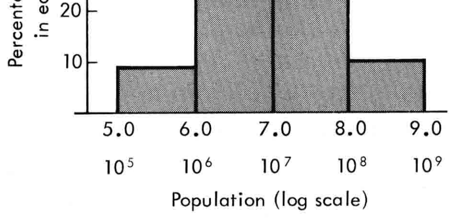


FIGURE 3-20 Logged vs. unlogged frequency distributions

Analysis (New York: Wiley, 1966); J. W. Richards, *Interpretation of Technical Data* (New York: Van Nostrand-Reinhold, 1967); and Joseph B. Kruskal, *op. cit.* For applications to political data see Hayward Alker and Bruce Russett, "Multifactor Explanations of Social Change," in Russett et al., *World Handbook of Political and Social Indicators* (New Haven, Conn.: Yale, 1964), 311–21.

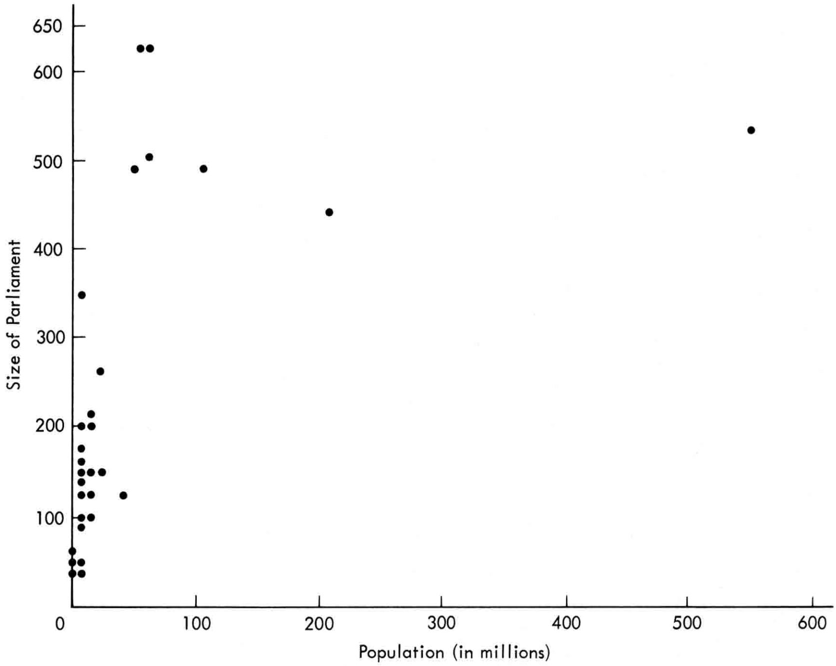


FIGURE **3-21a** Relationship between parliamentary size and population, 29 democracies—neither variable logged

Case IV is simply the usual two-variable regression model with both variables unlogged. We now consider the three cases in which at least one of the variables in the analysis is logged.

CASE I—BOTH THE DESCRIBING AND THE RESPONSE VARIABLE LOGGED

In the model

$$\log Y = \beta_1 \log X + \beta_0,$$

we estimate β_1 and β_0 by ordinary least squares by letting $X' = \log X$ and $Y' = \log Y$, which yields the linear form

$$Y' = \beta_1 X' + \beta_0.$$

How is the regression coefficient in the double-log case interpreted? Beginning with the regression

$$\log_{10} \mathrm{Y} = \beta_1 \log_{10} \mathrm{X} + \beta_0$$

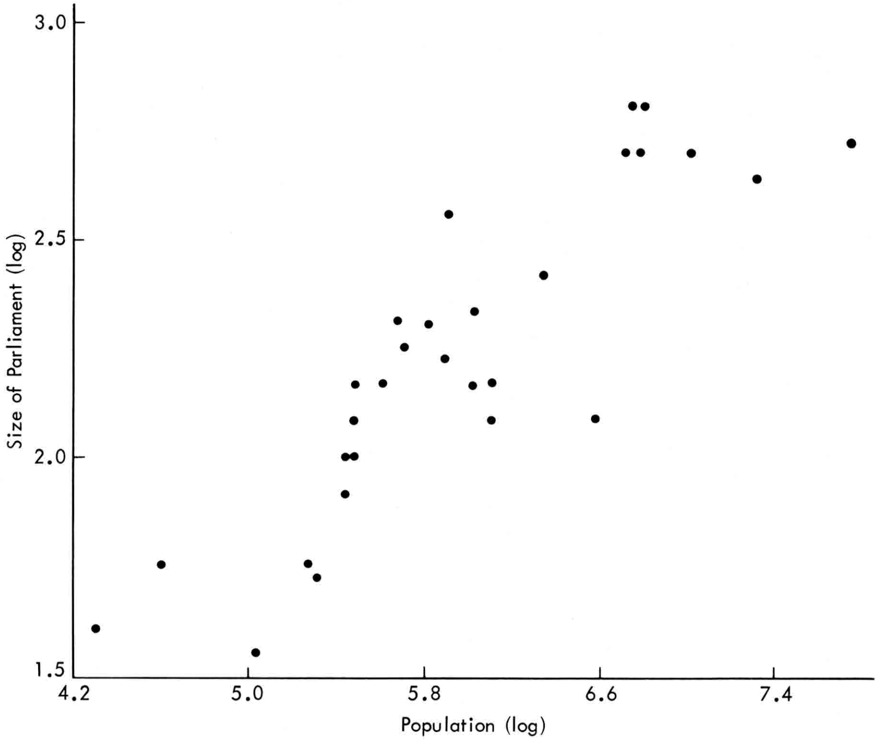


FIGURE 3-21b Relationship between parliamentary size (log) and population (log)—both variables logged.

and taking derivatives,

$$\frac{dY}{dX}\frac{1}{Y}\log_e 10 = \beta_1(\log_e 10)\frac{1}{X} + 0,$$

yields $\dfrac{dY}{dX}\dfrac{X}{Y} = \beta_1$

or $\beta_1 = \dfrac{dY/Y}{dX/X}$, which is the *elasticity* of Y with respect to X.

Thus β_1 measures the *percentage* change in Y with respect to a *percentage* change in X. The slope can be written approximately as

$$\beta_1 = \frac{\Delta Y/Y}{\Delta X/X}$$

and, when both the describing and the response variables are logged, the estimate of the slope assesses the proportionate change in Y resulting from a proportionate change in X. Note how this differs from the usual interpretation of the slope when both variables are unlogged (case IV):

$$\beta_1 = \frac{\Delta Y}{\Delta X}.$$

It is important to realize that fitting the model

$$\log Y = \beta_1 \log X + \beta_0,$$

does not *test* the assumption that there is, in fact, a proportionate relationship between X and Y. The logic is: *Assuming that there is a proportionate relationship between X and Y,* what is the best estimate of that proportionality or elasticity? Thus the regression answers the quantitative question by estimating a parameter in a model—on the assumption that the model is correct. We choose between competing models by comparing their goodness of fit, by thinking about their theoretical underpinnings, and by adding sufficient degrees of freedom in the model to allow the data to indicate the best fit. Our first example illustrates this point.

EXAMPLE 1 FOR THE LOG-LOG CASE: RELATIONSHIP BETWEEN PARLIAMENTARY SIZE AND POPULATION SIZE

Figure 3-22 shows the relationship, with both variables logged, between the population of a country and the size of its parliament for 135 countries of the world.[15] This relationship appears nearly linear in logarithms, and the fitted line is

$$\log_{10} \text{members} = .396 \log_{10} \text{population} - .564,$$

which explains, statistically at least, some 70.7 percent of the variation

[15] A discussion of the substantive issues involved in this relationship is found in Robert A. Dahl and Edward R. Tufte, *Size and Democracy* (Stanford, Calif.: Stanford University Press, 1973), Ch. 7.

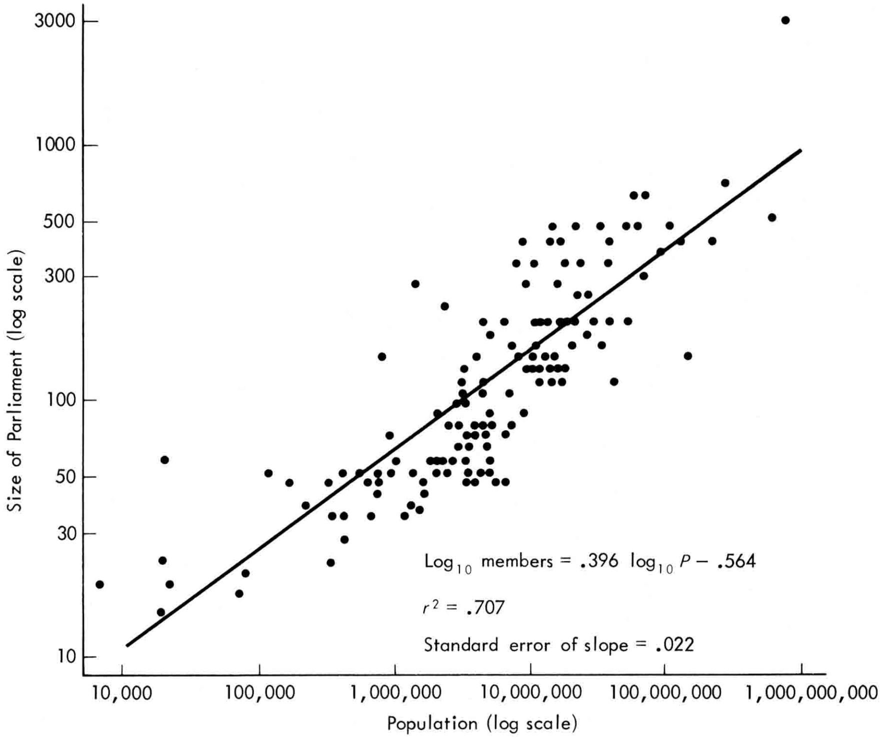


FIGURE **3-22** Population vs. parliament size—both variables logged

in parliamentary size. The estimated slope, .396, indicates that if a country was one percent above the average population of all countries, it was also typically about .4 percent above average with respect to size of parliament. A slightly more daring interpretation is to say that a change of one percent in population typically produces a change of .4 percent in parliamentary size.

Figure 3-22 and the residuals from the fitted line show a bend in the data—there is something of a threshold in the size of parliament for the smaller countries. For most of the countries with less than one million people, the observed points lie above the fitted line, indicating a tendency toward a minimum size of parliament around thirty members. We can improve upon the first fitted line for the 135 countries by examining some models that avoid the assumption of constant elasticity for all values of population (P) and take the bend in the data into account. One good approach, upon observing

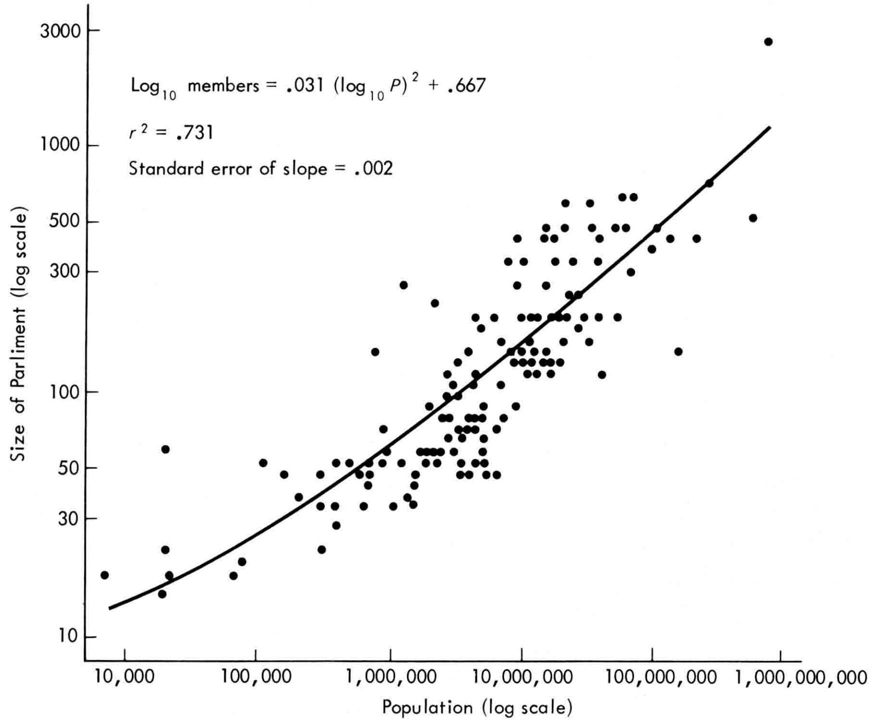


FIGURE 3-23 Fitted line with quadratic term

a curve in the data, is to introduce a quadratic term. The following fit, with its $(\log P)^2$ term, is our second model:

$$\log M = .031(\log P)^2 + .667.$$

Figure 3-23 shows the fit. This regression predicts 73.1 percent of the variation in the logarithm of parliamentary size—an improvement of 2.4 percentage points over the first model with no increase in the number of coefficients used in the model. What is the interpretation of this result? In particular, what does the regression coefficient mean? We get the answer by applying the same logic used in deriving the elasticity in the log-log case. The model is

$$\log_{10} M = \beta_0 + \beta_1(\log_{10} P)^2.$$

Taking derivatives, as before,

$$\frac{dM}{dP}\frac{1}{M}\log_e 10 = 2\beta_1(\log_e 10)(\log_{10} P)\frac{1}{P},$$

which yields

$$\frac{dM}{dP}\frac{P}{M} = \text{elasticity of } M \text{ with respect to } P$$

$$= 2\beta_1 \log_{10} P,$$

or, in our particular case,

$$= .062 \log_{10} P.$$

Thus in this model the elasticity of M with respect to P is a slowly increasing function of log P. For countries around 100,000, the elasticity of parliamentary size with respect to population is about .3; for countries of 100,000,000, it is nearly .5. Table 3-7 tabulates the relationship.

TABLE 3-7
Predictions of the Second Model

Population	*Log population*	*Elasticity of M with respect to* $P = .062 \log_{10} P$
10,000	4	.248
100,000	5	.310
1,000,000	6	.372
10,000,000	7	.434
100,000,000	8	.496
750,000,000	8.875	.550

The first model assumes that the elasticity is constant and provides an estimate under that untested assumption. The second model assumes that the elasticity varies as the population varies and provides an estimate under that untested assumption. The second is now favored because (1) visual inspection of the scatterplot and the residuals shows a bend in the data and (2) the second explains more variance than the first, even though both models estimate the same number of coefficients.

EXAMPLE 2 FOR THE LOG-LOG CASE: SIZE OF GOVERNMENTAL BUREAUCRACY AND POPULATION SIZE

For the fifty U.S. states, let B = the number of employees of the state government and let P = the number of people living in the state. Both P and B are highly skewed variables, and so we will work with log P and log B. Figure 3-24 shows log B plotted against log P.

Three sorts of general results could emerge from this analysis: (1) if a kind of Parkinson's Law held, then we would expect the bureaucracies of state governments to grow faster than the size of the state; (2) if there were, say, economies of scale, then we would expect bureaucracies to grow more slowly than the population of the state; and (3) the number of bureaucrats could grow in constant proportion to the size of the state. Obviously, other sorts of explanations can be used to explain the results of the analysis. The point here is that the number of employees of the state government can grow

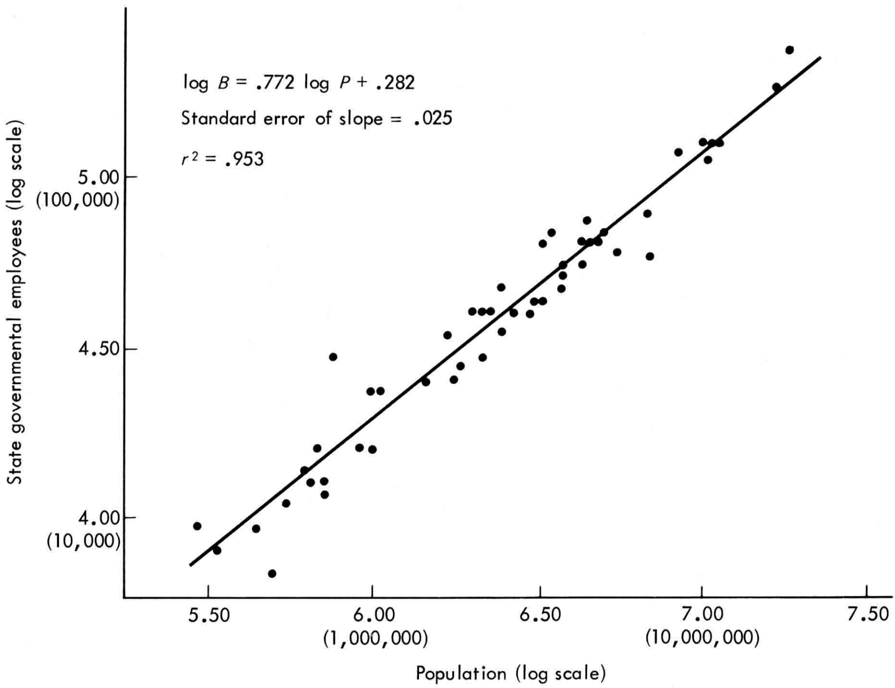


FIGURE 3-24 Population and state government employees

faster, slower, or at the same rate as the number of citizens in the state.

The model that helps to choose among these possibilities is

$$\log B = \beta_1 \log P + \beta_0$$

or letting $\beta_0 = \log\ c$ and taking antilogs puts the model in terms of the untransformed variables:

$$B = cP^{\beta_1}.$$

If β_1 is approximately one, then B approximately equals cP, which says that B grows linearly in direct proportion as P grows. In this case, there is support for what might loosely be called the "null hypothesis" concerning the relationship between size and the dependent variable. An example where β_1 would be very close to one and the null hypothesis accepted would be the relationship between the

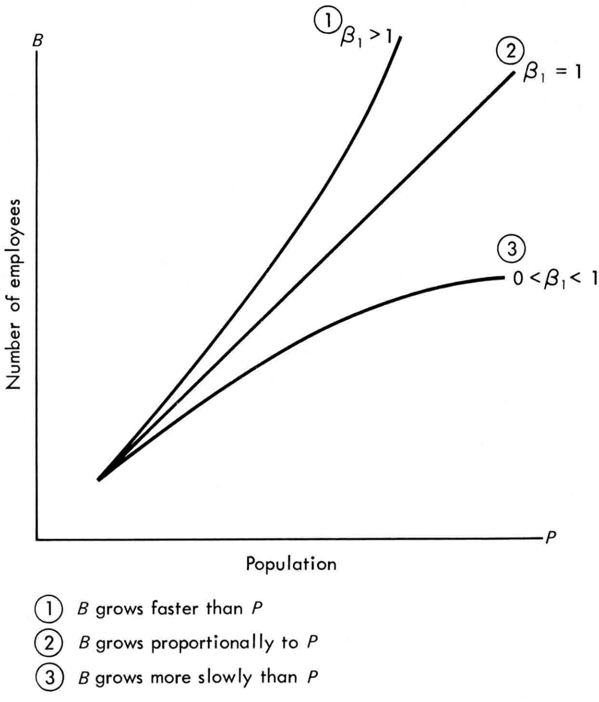


FIGURE 3-25 Three types of relationships between B and P

size of the population and the number of women in the population. In this case, given the sex ratio, c would be about .52.

In terms of the untransformed variables, if the estimated regression coefficient is greater than one, the slope increases as P increases. If β_1 lies between zero and one, the slope continually decreases. Figure 3-25 shows this result in a plot of the untransformed variables.

For the fifty states, we have the following results:

$$\log B = .772 \log P + .282,$$

$$\text{Elasticity} = \hat{\beta}_1 = .772,$$

$$\text{Standard error of elasticity} = .025, \qquad r^2 = .953.$$

Figure 3-24 shows the fitted curve.

The estimated elasticity is less than unity, indicating that the number of government employees grows somewhat more slowly than population. A change of one percent in the size of the population of a state is associated with a change of .772 percent in the number of government employees.

Note that the correlation coefficient is virtually useless in this problem. The square of the correlation provides a measure of the goodness of fit; but what is important is the estimate of the slope.

EXAMPLE 3 FOR THE LOG-LOG CASE: TESTING THE "CUBE LAW" RELATING SEATS AND VOTES WITH A LOGIT MODEL

One well-known description of the relationship between votes and seats in two-party systems is the "cube law."[16] The most economical statement of the law is that the cube of the vote odds equals the seat odds, where the vote odds are the ratio of the share of the votes received by one party divided by the share of the votes received by the competing party. For example, if both parties win 50 percent of the votes, then the odds are one to one. Figure 3-26 shows the line traced out by the cube law.

Quite a number of papers have touched upon the law and, in the last few years, the law has enjoyed a certain vogue and has been fitted to electoral outcomes in England, the United States, New Zealand, and, in a modified form, Canada. With one or two exceptions, discussions of the law are quite sympathetic, suggesting that it is

[16] This discussion follows E. R. Tufte, "The Relationship Between Seats and Votes in Two-Party Systems," *American Political Science Review*, 68 (June 1973), 540–54. Additional discussion of the paper is found in the *American Political Science Review*, 68 (March, 1974), 207–13.

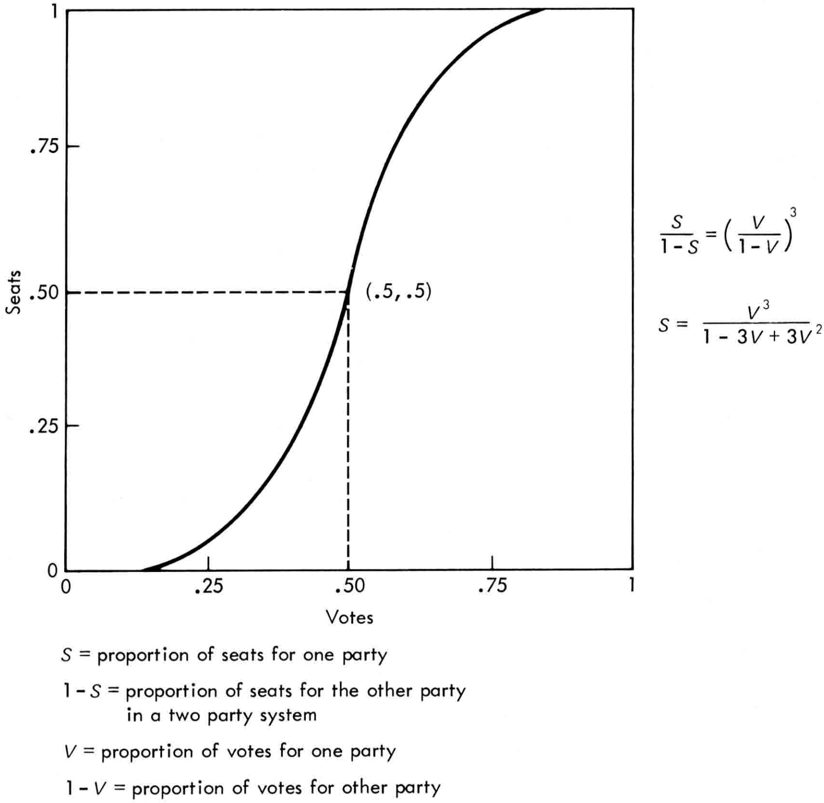


FIGURE **3–26** The cube law
SOURCE: Figure follows James G. March, "Party Representation as a Function of Election Results," *Public Opinion Quarterly*, 11 (Winter 1957–58), p. 524.

a useful and accurate description of electoral realities. Most studies consider no more than a few data points and conclude that the law fits rather well—although the quality of fit is usually assessed informally and no alternative fits are tried. Let us consider a direct test of the predictions of the cube law by using the log-log model. The law is

$$\frac{S}{1 - S} = \left(\frac{V}{1 - V}\right)^3.$$

The ratio of shares of seats and votes won by the two parties represents the odds that a party will win a seat or a vote. Taking logarithms yields

$$\log_e \frac{S}{1-S} = 3 \log_e \frac{V}{1-V},$$

and therefore in the regression of log-odds on seats against log-odds on votes,

$$\log_e \frac{S}{1-S} = \beta_0 + \beta_1 \log_e \frac{V}{1-V},$$

the cube law makes the simultaneous joint prediction that $\beta_0 = 0$ and $\beta_1 = 3$. Table 3-8 reports the results of tests of these predictions.

The table indicates that the cube law fits poorly in six of the seven trials. It fits quite well for the last eight elections in Great Britain, but otherwise its predictions are not confirmed. In short, it is not a "law." Since previous studies have not tested the exact joint predictions of the cube law (that is, $\beta_0 = 0$ and $\beta_1 = 3$) or used as extensive a collection of data, these results should be decisive in evaluating the empirical merits of the cube law.

Our previous analysis of seats and votes (Example 4) points to other defects in the cube law. The law hides important political issues because it implies that the translation of votes into seats is (1) unvarying over place and time, and (2) always "fair," in the sense that the curve traced out by the law passes through the point (50 percent votes, 50 percent seats), and the bias is zero.

As we have seen, these implications are not true. The rate of translation of votes into seats differs greatly across political systems, ranging between gains of 1.3 to 3.7 percent in seats for each 1.0 percent gain in votes. Also the results in Table 3-8 indicate that some electoral systems persistently favor a particular party; the votes-seats curve traced out by the data does not inevitably pass close by the point (50 percent votes, 50 percent seats).

The model estimated in the test of the cube law is called a "logit model." Define the odds in favor of a party winning a seat as $S/(1 - S)$ and the vote odds as $V/(1 - V)$. The logit model is the regression of the logarithm of seat odds against the logarithm of vote odds (a regression used earlier to test the specific predictions of the cube law):

TABLE 3-8
Testing the Predictions of the Cube Law (and Simultaneously Estimating the Logit Model)

	$\hat{\beta}_0$	$\hat{\beta}_1$	*Standard error of slope*	r^2	*Does $\beta_0 = 0$ and $\beta_1 = 3$ as cube law predicts?*	*Is $\beta_0 \neq 0$; that is, is there a significant bias?*
Great Britain	−.02	2.88	.30	.94	Yes	No bias
New Zealand	−.12	2.31	.27	.91	No	Yes, there is a bias
United States, 1868–1970	.09	2.52	.24	.68	No	Yes
United States, 1900–1970	.17	2.20	.15	.86	No	Yes
Michigan	−.17	2.19	.43	.76	No	Yes
New Jersey	−.77	2.09	.59	.29	No	Yes
New York	−.23	1.33	.19	.74	No	Yes

$$\log_e \frac{S}{1 - S} = \beta_0 + \beta_1 \log_e \frac{V}{1 - V}.$$

Since both variables are logged, the estimate of the slope, $\hat{\beta}_1$, is the estimated elasticity of seat odds with respect to vote odds; that is, a change of one percent in the vote odds is associated with a change of $\hat{\beta}_1$ percent in seat odds.

The logit model has the advantage over the linear fit used in Example 4 of producing a reasonable predicted value for the share of seats for all logically possible values of the share of votes; the predicted values stay between 0 and 100 percent seats for any percentage share of votes. As noted earlier, this is only a theoretical virtue, since the more extreme values do not occur empirically. The logit model also provides a direct test of the hypothesis that an electoral system is unbiased, since $\beta_0 = 0$ in an unbiased system. As shown in Table 3-8, there is a statistically significant bias in all cases except Great Britain.

CASE II—RESPONSE VARIABLE LOGGED, DESCRIBING VARIABLE NOT LOGGED

Here we have the model of the form

$$\log Y = \beta_0 + \beta_1 X.$$

One particularly interesting application of such a model derives from the exponential:

$$Y = ae^{bX}.$$

Taking natural logarithms and letting $c = \log_e a$ puts this model into the form of case II:

$$\log_e Y = c + bX.$$

This exponential model can be estimated by ordinary least squares, and the regression coefficient has the following interpretation:

> In the model $Y = ae^{bX}$, $b \times 100$ is approximately equal to the *percent increase in Y per unit increase in X,* if b is small (say, less than .25).

The proof of this statement relies on the series expansion of e^X:

Percent increase in Y per unit increase in X

$$= \frac{\dfrac{\Delta Y}{Y}}{\Delta X}$$

$$= \frac{Y_2 - Y_1}{Y_1} \qquad (\text{since } \Delta X = X_2 - X_1 = 1)$$

$$= \frac{ae^{bX_2} - ae^{bX_1}}{ae^{bX_1}}$$

$$= e^{(bX_2 - bX_1)} - 1$$

$$= e^b - 1 \quad (\text{since } X_2 - X_1 = 1)$$

$$= [1 + b + \frac{1}{2!}b^2 + \frac{1}{3!}b^3 + \ldots] - 1,$$

by the expansion of e^b. So, if b is small, we can drop the higher-order terms, leaving

$$\approx (1 + b) - 1 = b.$$

Thus $b \times 100$ equals the percent increase in Y associated with a unit increase in X.[17]

The logarithm of the response variable is used in estimating rates of increase over time. Table 3-9 shows the gross national product of Japan from 1961 to 1970. Note the increasing absolute increase in GNP growth—GNP (the yearly absolute increase) itself increases over time. One process generating such increasing increases is a constant *percentage* growth rate—just like compound interest. What is the appropriate model for a constant percentage growth rate? Consider compound interest, at i percent per year. Beginning the first year with principal P_0 leads to principal P_t after t years:

$$P_t = P_0 (1 + i)^t.$$

For example, after one year:

$$P_1 = P_0(1 + i).$$

After two years

$$\begin{aligned} P_2 &= P_1(1 + i) \\ &= P_0(1 + i)^2, \end{aligned}$$

and so on. To put this into slightly more familiar notation:

$$Y_t = Y_0(1 + i)^t.$$

Taking the logarithm of both sides

$$\log Y_t = \log[Y_0(1 + i)^t],$$

$$\log Y_t = \log Y_0 + \log(1 + i)^t,$$

$$\log Y_t = \log Y_0 + t \log(1 + i).$$

[17] An application of this interpretation is found in Philip E. Sartwell and Charles Anello, "Trends in Mortality from Thromboembolic Disorders," in Advisory Committee on Obstetrics and Gynecology, Food and Drug Administration, *Second Report on the Oral Contraceptives* (Washington, D.C.: U.S. Government Printing Office, 1969), 37–39.

Now let

$$\beta_0 = \log Y_0,$$

$$\beta_1 = \log(1 + i),$$

and we have the model

$$\log Y_t = \beta_0 + \beta_1 t$$

—that is, case II. The model is estimated by letting $Y = \log Y_t$, yielding

$$Y = \beta_0 + \beta_1 t,$$

the usual linear model.

Figure 3-27 shows the GNP of Japan plotted on both an absolute scale and a logarithmic scale. Note how, for these data, the log scale throws the data points into a straight line. The changes in the logarithm of GNP are relatively constant (Table 3-9), indicating a relatively constant percentage rate of growth over time. The line for log GNP fits considerably better than the line for absolute GNP—as the r^2 shows. The fitted line for the logarithmic case is

$$\log_{10} \text{GNP} = 1.627 + .064\,t.$$

The rate of growth, i, can be estimated by going back to the original linearization of the model,

$$\beta_1 = \log(1 + i),$$

and solving by taking antilogarithms. This yields

$$\hat{i} = .159,$$

or a growth rate of almost 16 percent per year.[18]

This is the yearly rate of growth. An instantaneous rate of growth can be estimated by fitting the model

[18] Unfortunately the estimate, $\hat{i}$, is biased. It does not have least-squares properties because the sum of squares was minimized with respect to log GNP rather than GNP over time.

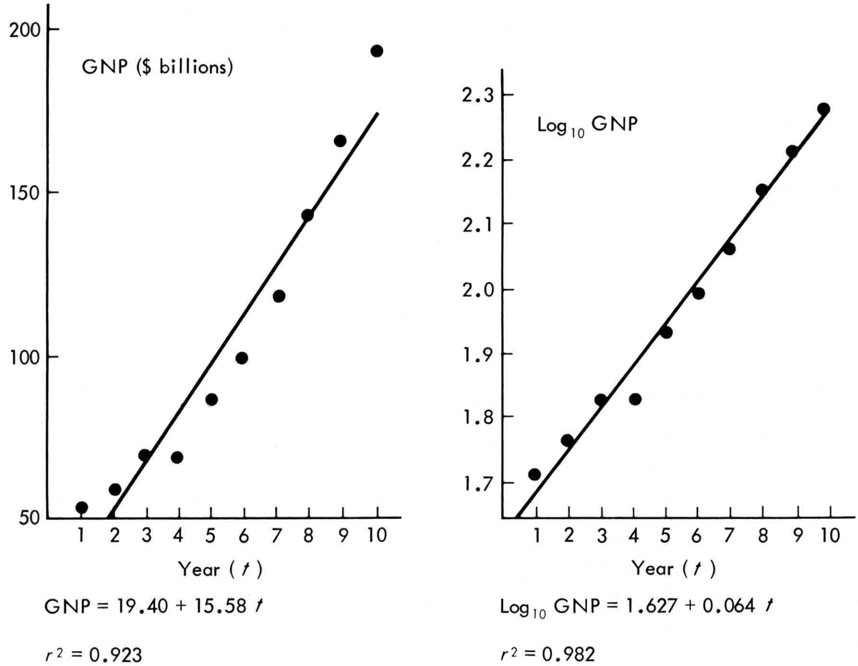


FIGURE 3-27 Growth of GNP, Japan, 1961–1970

$$\log_e Y = \beta_0 + \beta_1 t.$$

Differentiating gives

$$\beta_1 = \frac{dY/Y}{dt},$$

the percentage rate of growth in Y.

Finally, a growth rate can be estimated quite soundly without the regression model, simply by taking the average (mean, median, or midmean) of the yearly growth rates, or the average of the logarithm.

CASE III—RESPONSE VARIABLE UNLOGGED, DESCRIBING VARIABLE LOGGED

The model is

$$Y = \beta_0 + \beta_1 \log X.$$

TABLE 3-9
Gross National Product, Japan, 1961–1970

Year	*t*	*GNP ($ Billion)*	*Yearly increase in GNP*	*$\log_{10}$ GNP*	*Yearly increase in $\log_{10}$ GNP*
1961	1	53		1.72	
			6		.05
1962	2	59		1.77	
			9		.06
1963	3	68		1.83	
			0		.00
1964	4	68		1.83	
			17		.10
1965	5	85		1.93	
			12		.06
1966	6	97		1.99	
			19		.07
1967	7	116		2.06	
			26		.09
1968	8	142		2.15	
			24		.07
1969	9	166		2.22	
			31		.08
1970	10	197		2.30	

If the logarithm of the describing variable is taken to the base 10, the regression indicates that a change in the order of magnitude of X—that is, a tenfold increase in X—is associated with a change of β_1 units in Y.

Sometimes it is useful to take the logarithm to the base 2 in this model. In such a case, the regression coefficient estimates the increase in Y when X doubles. And so when X is measured with respect to time, the estimate of the regression coefficient may be said to assess the "doubling time" of Y with respect to X. It is easy to prove that when X doubles, Y increases by β_1 units. The model is

$$Y = \beta_0 + \beta_1 \log_2 X.$$

Now suppose X doubles:

$$\begin{aligned} Y_{\text{new}} &= \beta_0 + \beta_1 \log_2 2X \\ &= \beta_0 + \beta_1(\log_2 2 + \log_2 X) \\ &= \beta_0 + \beta_1 \log_2 X + \beta_1 \\ &= Y + \beta_1 \end{aligned}$$

—that is, the value of Y after X doubles is the old value of Y plus β_1. Thus Y increases by β_1 units when X doubles.

Consider the following application of this model. Kelley and Mirer have developed a rule predicting how voters will vote; the predictions are made on the basis of an interview with the voter D days before the election. After the election, the voter is reinterviewed and asked how he or she voted. Thus it is possible to find the rate of error in prediction—and such errors might well be related to how many days before the election the voter was interviewed. If D were 1000 days, to take an extreme example, the error rate in prediction would be higher than if D were one day. The researchers analyzed the data first with a linear model, then with a logarithmic model:

> A simple linear regression of the first of these variables on the second shows them to be strongly related. The equation yielded is:
>
> $$\text{rate of error} = 17.4 + .23(\text{days before election}).$$
>
> In a statistical sense this relationship explains some 28 percent of the variance in the dependent variable, and, since the standard error of the estimated coefficient is .07, the relationship is statistically significant ($t = 3.15$). Most interesting, perhaps, is the implication of the equation's constant term: Had the interviews of these respondents been conducted on election day, the mean rate of error in predicting their votes would have been 17.4 percent. . . .
>
> And it is quite possible that this value for the constant term is too high. The volume of partisan propaganda is normally much heavier in the last two or three weeks of a presidential campaign than it is earlier. We might therefore suppose the relationship between time and changes of opinion to be like that shown in Figure [3-28], in which the likelihood of such changes (and thus the error rates of our predictions) at first increases rapidly with increases in the number of days between election day and the time the opinions were expressed, then more slowly. By regressing the rates of error in our predictions for groups of respondents on the logarithm (to the base 2) of the mean number of days before election day that the respondents in each group were interviewed, one can see if a curve like that shown in Figure [3-28] fits the data that entered into the first regression. The equation produced by this new regression is:
>
> $$\text{rate of error} = 5.3 + 4.03(\log_2 \text{days before election}).$$
>
> This second equation accounts for as much of the variance in the dependent variable as did the first and yields an equally reliable estimate of the regression coefficient ($r^2 = .28$, $t = 3.14$). The value of the equation's constant term implies that our mean rate of error in predicting the votes of groups of respondents would have been 5.3 percent . . . if those respondents had been interviewed one day

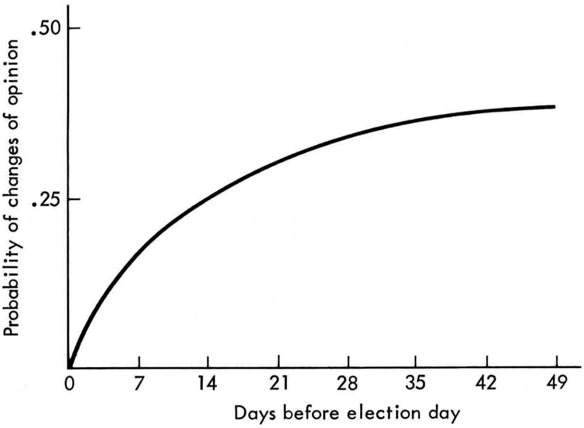


FIGURE 3-28 Hypothetical relationship between the likelihood that opinions will change and the time that attitudes toward parties and candidates are expressed

before election day. The equation as a whole implies that, starting from the day before the election, the error rate in predictions derived from the Rule will rise by four percentage points with each doubling of the length of time before election day that respondents are interviewed.[19]

Example 7: Regressions Aren't Enough—Looking at the Scatterplot

F. J. Anscombe has constructed a nice set of numbers illustrating why it is important to look at scatterplots along with the fitted equation.[20] Table 3-10 shows four sets of data. Their remarkable property is that all four yield exactly the same result when a linear model is fitted. The regression in all four cases is:

$$Y = 3.0 + .5X,$$

$$r^2 = .667, \text{ estimated standard error of } \beta_1 = 0.118,$$

[19] Stanley Kelley, Jr., and Thad W. Mirer, "The Simple Act of Voting," *American Political Science Review*, 68 (June 1974), pp. 582–83.

[20] F. J. Anscombe, "Graphs in Statistical Analysis," *American Statistician*, 27 (February 1973), 17–21. Copyright 1973 by the American Statistical Association. Reprinted by permission.

TABLE 3-10
Four Data Sets

DATA SET 1		DATA SET 2	
X	Y	X	Y
10.0	8.04	10.0	9.14
8.0	6.95	8.0	8.14
13.0	7.58	13.0	8.74
9.0	8.81	9.0	8.77
11.0	8.33	11.0	9.26
14.0	9.96	14.0	8.10
6.0	7.24	6.0	6.13
4.0	4.26	4.0	3.10
12.0	10.84	12.0	9.13
7.0	4.82	7.0	7.26
5.0	5.68	5.0	4.74
DATA SET 3		DATA SET 4	
X	Y	X	Y
10.0	7.46	8.0	6.58
8.0	6.77	8.0	5.76
13.0	12.74	8.0	7.71
9.0	7.11	8.0	8.84
11.0	7.81	8.0	8.47
14.0	8.84	8.0	7.04
6.0	6.08	8.0	5.25
4.0	5.39	19.0	12.50
12.0	8.15	8.0	5.56
7.0	6.42	8.0	7.91
5.0	5.73	8.0	6.89

SOURCE: F. J. Anscombe, *op. cit.*

mean of $X = 9.0$,

mean of $Y = 7.5$, for all four data sets.

And yet the four situations—although numerically equivalent in major respects—are substantively very different. Figure 3-29 shows how very different the four data sets actually are.

Anscombe has emphasized the importance of visual displays in statistical analysis:

> Most textbooks on statistical methods, and most statistical computer programs, pay too little attention to graphs. Few of us escape being indoctrinated with these notions:

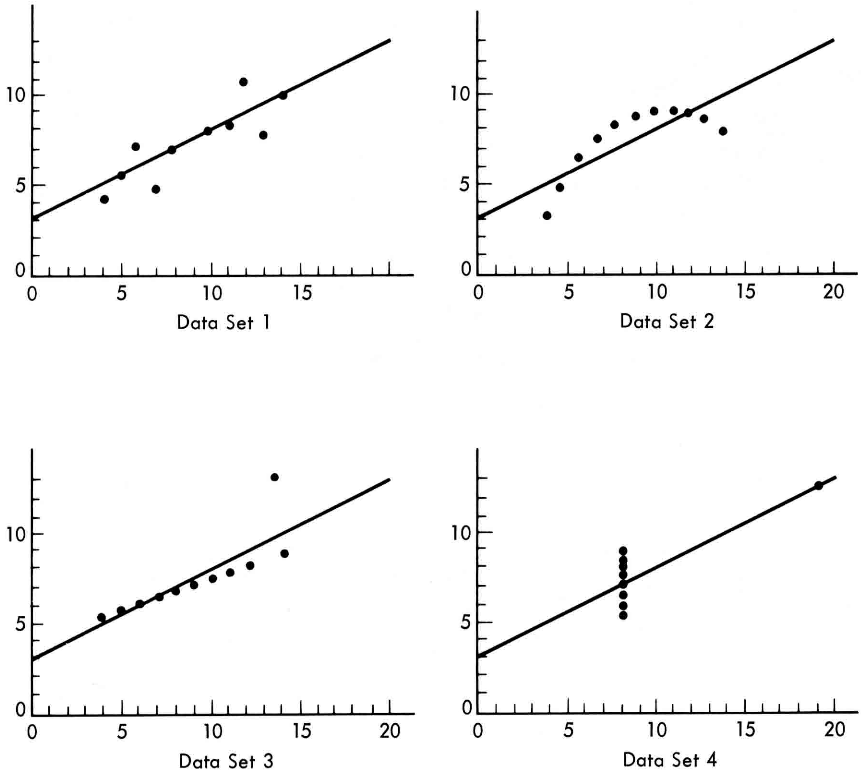


FIGURE **3-29** Scatterplots for the four data sets of Table 3-10
SOURCE: F. J. Anscombe, *op cit.*

(1) numerical calculations are exact, but graphs are rough;
(2) for any particular kind of statistical data there is just one set of calculations constituting a correct statistical analysis;
(3) performing intricate calculations is virtuous, whereas actually looking at the data is cheating.

A computer should make *both* calculations *and* graphs. Both sorts of output should be studied; each will contribute to understanding.

Graphs can have various purposes, such as: (i) to help us perceive and appreciate some broad features of the data, (ii) to let us look behind those broad features and see what else is there. Most kinds of statistical calculation rest on assumptions about the behavior of the data. Those assumptions may be false, and then the calculations may be misleading. We ought always to try to check whether the assumptions are reasonably correct; and if they are wrong we ought to be able to perceive in what ways they are wrong. Graphs are very valuable for these purposes.[21]

Up until now we have considered only one-variable explanations of the response variable. But the world is surely often more complicated than that and response variables have more than a single cause. In the next chapter, we examine the *multiple regression* model which allows us to take into account effectively several explanatory variables—at least some of the time.

[21] Anscombe, *op. cit.*, p. 17.