IMPROVING DATA ANALYSIS IN POLITICAL SCIENCE

By EDWARD R. TUFTE*

I. The Problem

STUDENTS of politics use statistical and quantitative techniques to: *summarize* a large body of numbers into a small collection of typical values;

confirm (and perhaps sanctify) the results of the analysis by using tests of statistical significance that help protect against sampling and measurement error;

discover what's going on in their data and expose some new relationships; and

inform their audience what's going on in the data.[1]

Most textbooks of social statistics tell us a lot about summarizing and sanctifying in their many chapters on descriptive statistics and significance tests, but they have little to say about how to discover the unanticipated or how to learn something new from the data. Graphical techniques are, for example, among the most powerful procedures for both discovering and informing. Yet most textbooks give only the most humdrum, and often misleading, advice about displaying the data—such as the "necessity" for showing the zero point and for equal class intervals. With this kind of advice, students naturally come away with the impression that the only point of graphing is to make a pie diagram showing how next year's budget is going to be spent. Similarly, most textbook discussions of how to fit straight lines to data—potentially the most powerful technique of data analysis available in many situations—are devoted to explaining how to see whether a particular coefficient is significantly different from zero (sanctification again). Little is said about how to analyze residuals or how to transform the variables entering a regression—techniques that will help us discover patterns in the data far better than any significance test. Finally, almost to guar-

* I wish to thank Hayward Alker, Jr., Stanley Kelley, Jr., Gerald Kramer, John McCarthy, Susanne Mueller, Walter Murphy, Dennis Thompson, and John Tukey for their advice and criticism. I also thank Joseph Verbalis, who constructed the figures.

[1] For similar categories, see John W. Tukey and M. B. Wilk, "Data Analysis: Techniques and Approaches," *Proceedings of the Symposium on Information Processing in Sight Sensory Systems* (Pasadena, California Institute of Technology, November 1965). This paper is also reprinted in Edward R. Tufte, ed., *The Quantitative Analysis of Social Problems* (Reading, Mass. 1969).

antee that budding social scientists learn little about discovering and informing, most textbooks of social statistics contain few, if any, actual case histories of data analysis.[2] No wonder most questions concerning data analysis ask whether a scale is "really" an interval scale or whether a one-tailed t-test is appropriate in a particular situation.

The results of all this bad advice show up in many recent quantitative efforts in political science. For example, consider the essays collected in *Quantitative International Politics*, a book representative of many quantitative studies of politics.[3] Instead of residuals off a regression line, we see almost every page littered with the results of significance tests often complete with all the relentless detail of computations, degrees of freedom, test statistics, and probability levels. Residuals (that part of the variation that is "unexplained") are not analyzed in any of the essays simply because no one fitted a line to any data. And even though there is lots of talk about how complicated the world is, no author uses multiple regression, which would allow the inclusion of more than a single explanatory variable. Instead of a sensitive analysis of the problem of multicollinearity (the difficulty in separating out the independent effects of highly correlated describing variables) and its immediate and profound consequences for testing theories of politics, we see a single correlation matrix twelve pages long in an article that totals twenty-eight pages.

In short, then, much of current political data analysis leaves the impression that the important uses of quantitative methods in the study of politics are either to summarize a large body of data or to sanctify an observed relationship at the .05 level of significance.

How can we do better?

From the point of view of discovering and informing, the following four points help summarize a good approach toward effective data analysis:

(1) Significance tests deserve a secondary role in data analysis. They can be useful, but only after we look at the right things. Significance tests cannot tell us what to think.

(2) The distinction between interval and ordinal measurement is usually of little importance in data analysis. The wise assignment of numbers to ordered categories, coupled with the use of techniques that exploit the properties of numbers, is generally preferable to working with ordered categories.

[2] An important exception is W. Allen Wallis and Harry V. Roberts, *Statistics: A New Approach* (Glencoe, Ill. 1956).

[3] J. David Singer, ed. (New York 1968).

(3) Looking at correlations is only a partial, first step in the analysis. Correlations are often misleading.

(4) The most effective method of data analysis usually begins by fitting lines to relationships between variables (transformed variables if necessary) and then continues by examining, with the aid of graphs and scatterplots, deviations off the fitted line.

Many people have argued one way or another about the first two points above, and these issues, broadly speaking, are not settled. But if we consider the issue of the utility of significance tests and the problem of levels of measurement from the point of view of discovering and informing—in other words, from the point of view of the pragmatic data analyst—we should find propositions 1 and 2 fairly easy to accept. Let us now look at the four basic points in more detail.

II. The Reconstructed Science of Significance Testing

The overemphasis on significance testing has arisen in part, I think, because some have taken the reconstruction of what scientists do more seriously than actual scientific practice.[4] Thus the rote paradigm of significance testing—Assumptions (Level of Measurement, Model, and Hypothesis); Sampling Distribution; Significance Level and Critical Region; Computing the Test Statistic; and, behold, The Decision—represents a severe and impractical formalization that fails to provide useful guides to effective data analysis. Such a paradigm often serves only to make researchers feel guilty that they have violated one of the unrealistic assumptions of statistical significance tests.

What good are tests of significance, then? Such tests help protect against the possibility that a relationship arises because of happenstance in a random sample.[5] They are also useful as a rough sort of screening device in the analysis of data collected nonrandomly. They may help us adjust our feeling of certainty or uncertainty about a result. Thus the responsible investigator will use tests against the null hypothesis or, often better, confidence intervals to assess the stability of his results. Significance levels are often misused, however, because the dichotomy between "significant" and "nonsignificant" is too sharply drawn and the investigator regards those relationships that reach the .05 level (sometimes the .01 level is the sacred probability) as being the only truly meaningful results. This is bad practice; the relevance of a result does

[4] See Abraham Kaplan, *The Conduct of Inquiry* (San Francisco 1964), chap. 1.

[5] Leslie Kish, "Some Statistical Problems in Research Design," *American Sociological Review*, 24 (June 1959), 336. Another good discussion of significance tests is William H. Kruskal, "Tests of Significance," *International Encyclopedia of the Social Sciences* (New York 1968), vol. 14, 238-50.

not hinge on its exact significance level. Emphasizing the importance of substantive judgment in interpreting the results of data analysis, one statistician has proposed the "interocular trauma test" of significance: you know what the data mean when the conclusion hits you between the eyes. Edwards, Lindman, and Savage comment further that the "enthusiast's interocular trauma may be the skeptic's random error. A little arithmetic to verify the extent of the trauma can yield great peace of mind for little cost."[6]

Finally, probability levels and test statistics tell us very little about the strength and nothing about the substantive significance of a relationship. The important question is, "Does the result show a relationship which is of substantive interest because of its nature and magnitude?"[7] Significance tests are silent on this matter. Mosteller and Hammel throw light on the problem by noting that when the investigator uses tests of statistical significance, he then faces "the standard difficulty of what to do with the relationships once significance is established. Significance tests have little to say about this. What to do is never made entirely clear, and, as far as we can see, it will be up to the investigator to use his own judgment in appraising and weighting reports that show relations with conditions of observation. Perhaps it will never be possible to improve on individual judgment in such matters; the treatment may be an open statistical problem, or it may be that regression methods could aid the research worker."[8]

In short, then, it seems obvious to suggest that scientific, mathematical, or statistical jargon should not be used merely to sanctify the results of political analysis. Hempel has put the matter clearly: "To say that 'A man *M* walks down a street *S*' does not increase the scientific validity of a statement." And to attach a test statistic and a statistical significance level to a statement does not help it much either.

III. Getting out of the Rut of Cross-tabulations: Pinning Numbers on Ordered Categories

How seriously should a data analyst take the distinction between ordinal and interval measurement? Many have taken the distinction very, very seriously—almost to the point of paralysis. While there is

[6] Ward Edwards, Harold Lindman, and Leonard J. Savage, "Bayesian Statistical Inference for Psychological Research," *Psychological Review*, 70 (May 1963), 217.

[7] Kish, 336.

[8] Frederick Mosteller and E. A. Hammel, book review, *Journal of the American Statistical Association*, 58 (September 1963), 836.

a conceptual distinction between ordinal and interval measurement,[9] the issue that should concern the data analyst is whether this distinction has any practical meaning for his work. One good reason for pinning numbers on ordered categories is that the researcher often knows more about the phenomenon than the mere ordering of observations implies; thus, assigning numbers helps to build that additional information into measurement. For example, if the researcher knows that $A>B>C>D$ and, on the basis of his substantive understanding of the thing being measured, he also knows that D is far from C compared to, say, the difference between A and B, then the measurement should incorporate this information by assigning, for example, the numbers 1, 2, 3, and 9 to the ordered categories. Of course it is arbitrary.[10] The point, as Tukey has put it, is to be *wisely* arbitrary. The argument raises two issues:

(1) Should we pin numbers on ordered categories?

(2) If yes, just exactly *what* numbers are to be assigned to each category?

From the point of view of doing the detective work of data analysis, we want to assign numbers to orderings. By doing this, we improve measurement by taking advantage of any additional information above and beyond the fact of ordering. Numbers put more substance into measurement. An additional gain is that considerably more powerful techniques can be used in the analysis of the data and thereby increase our chances of learning something new. In the face of these potential gains—better measurement and better analysis—there appears to be very little in the way of potential costs. By sticking rigidly to the distinction between ordinal and interval levels of measurement (a distinction often difficult to make in practice), we are, in effect, censoring the data and shutting off part of what the data could tell us—if we would only let them.[11]

Admitting, then, the possibility that numbers actually should be assigned to ordered categories, just exactly what numbers should be

[9] For a recent statement of S. S. Stevens, see his "Measurement, Statistics, and the Schemapiric View," *Science*, 161 (August 30, 1968), 849-56.

[10] For a discussion of the problem of "being arbitrary," see J. C. Nunnally, *Psychometric Theory* (New York 1967), chap. 1.

[11] One common practice is to convert the numerical values of variables into ordered ranks before computing measures of association. Such a transformation, presumably made because it somehow seems statistically more conservative (it is not), may throw away useful information in the data and also sometimes discourage efforts at multivariate analysis. One other alternative is to employ some of the nonmetric multivariate methods.

assigned? Several methods have been proposed,[12] but three fairly useful and simple rules are:

(1) The assignment should incorporate the investigator's substantive understanding of the thing being measured.

(2) The simple linear assignment of numbers to categories (e.g., assigning 1, 2, 3, 4 to four ordered categories) usually won't do. Such a linear assignment is not, in any way, a sounder or more conservative choice than any other assignment. And the chances are that such an assignment is not consistent with the first principle of incorporating substantive information into the measurement.

(3) The assignment should often be made so that the distribution of counts begins to look somewhat like a normal distribution.

If the concern is, for example, the magnitude of internal conflict in a society, the measurement might begin with the four ordered categories: (a) petitions against the government, (b) peaceful, massive demonstrations, (c) violent demonstrations, and (d) civil war. Plainly, the category (d) of civil war is in a league by itself so far as internal conflict is concerned. The assignment of numbers to the four categories should take this into account. One such assignment would be the numbers 1, 2, 5, and 17. It would be even better to start with more than four categories—in fact, with as many categories as possible—especially since there is no reason to believe that internal conflict in societies is distributed in four clumps (and certainly not in four equally spaced clumps, as the assignment of 1, 2, 3, and 4 would assume). But, at any rate, we should never work with only two categories. Throwing everything into two bins by dichotomizing the data is just about the most severe form of censorship we could impose on the data. Even with the choice of the optimal cutting point, a major amount of information is lost by dichotomizing the data.

IV. The Pitfalls of Correlations

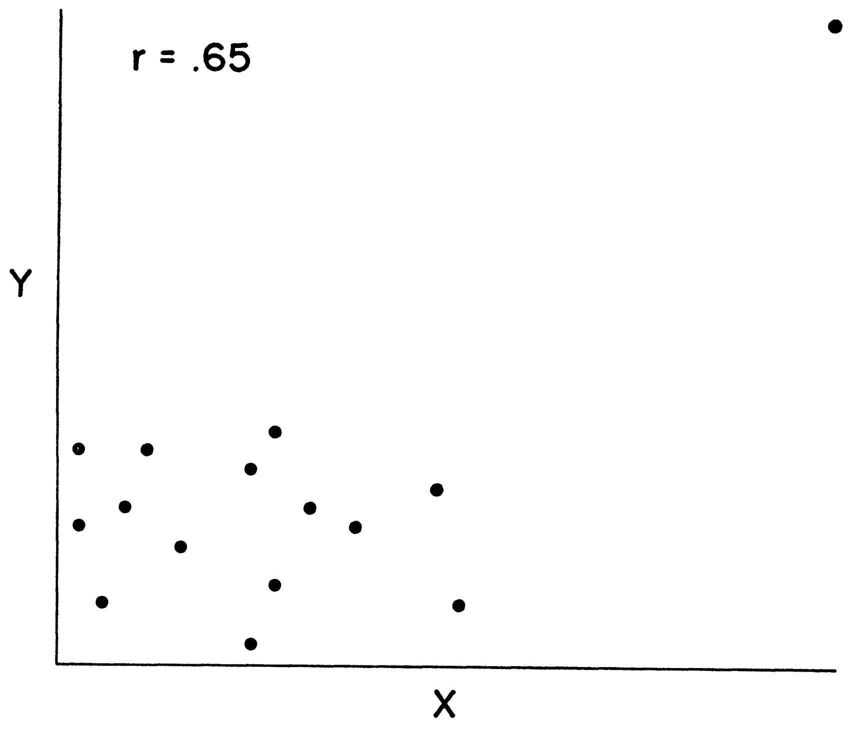
The correlation coefficient is often used to summarize the relationship between two variables. It is overworked. The contributors to *Quantitative International Politics* report a grand total of about 1600 correlations in their various essays. Or, taking another example,

[12] The discussion here is necessarily rather brief. For more information, see Robert P. Abelson and John W. Tukey, "Efficient Conversion of Non-Metric Information into Metric Information," *Proceedings of the Social Statistics Section of the American Statistical Association* (Washington 1959), 226-30 (also in Tufte); Abelson and Tukey, "Efficient Utilization of Nonnumerical Information in Quantitative Analysis: General Theory and the Case of Simple Order," *Annals of Mathematical Statistics*, 34 (December 1963), 1347-69; and Roger N. Shepard, "Metric Structures in Ordinal Data," *Journal of Mathematical Psychology*, 3 (1966), 287-315.

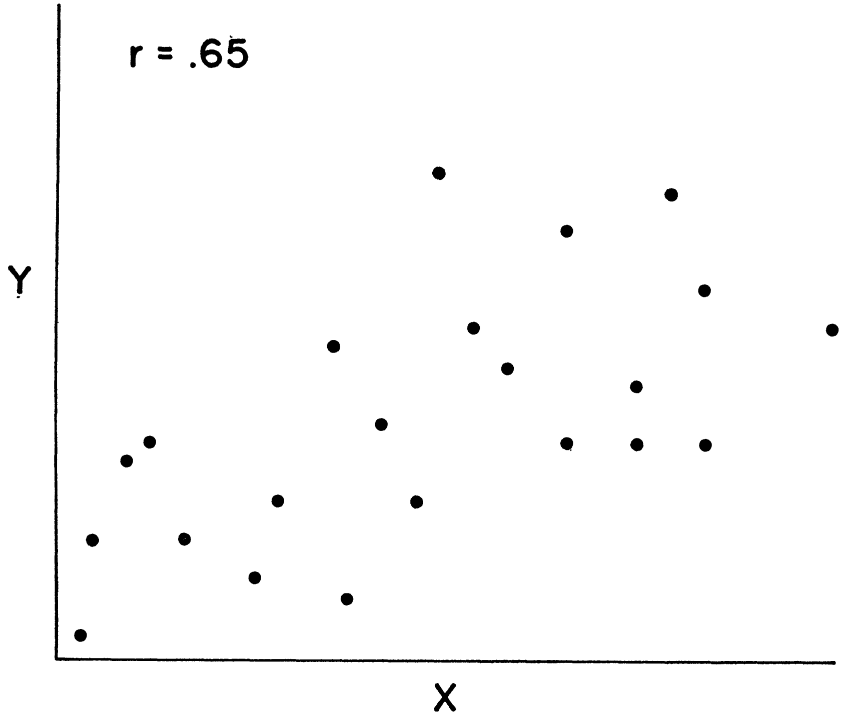
Thomas Dye's *Politics, Economics, and the Public* includes more than 5000 correlations. Such coefficients have serious defects; indeed, their faults are often so great that some have recommended that "most correlation coefficients should never be calculated."[13] What are the pitfalls involved in the use of correlations? What are the alternatives to correlation coefficients?

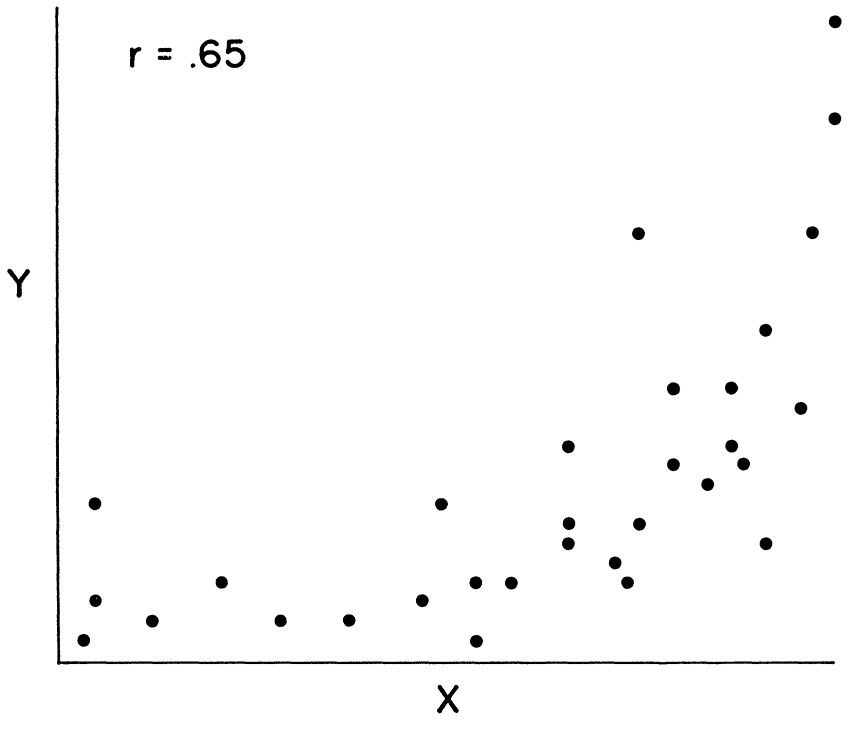
The three scatterplots shown below convey widely divergent information about the relationship between X and Y. The first plot shows no relationship, discounting the extreme outlier on both measures. The second plot suggests a moderately strong linear relationship between X and Y. The third plot reveals a rather marked curvilinear relationship between X and Y, indicating 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.*

This example shows that a correlation coefficient is a pretty poor way of summarizing the actual data as revealed in the scatterplot. One practical moral here is that any computer program that produces a correlation matrix should also produce scatterplots of the relationships. A



[13] John Tukey, "Causation, Regression, and Path Analysis," in Oscar Kempthorne and others, eds., *Statistics and Mathematics in Biology* (Ames, Iowa 1954), 38.


r = .65
Y
X


r = .65
Y
X

more obvious moral is, of course, that people should use scatterplots more than they do. A major benefit of scatterplots and of most graphical procedures is that they allow the reader to decide for himself how much he wants to learn from the data—instead of having the relationship summarized, and perhaps maimed, by a correlation coefficient or other summary statistic.

Sometimes correlations take us too far away from the substantive matters at hand. For example, consider some hypothetical data relating the number of dollars spent for precinct work on election day and the voter turnout in each precinct. One way to summarize such data is to say, "The correlation between dollars and votes is 0.73." That doesn't say as much as does the statement: "On the average, every $10 produces 1.7 votes." Expressing the result as a correlation coefficient is less informative in this case than is expressing the result as the slope of the line fitting the data. Often both coefficients are informative.

The faults of correlation coefficients do not disappear when correlations are built into a more complex model through the use of partial correlations (as is frequently done in efforts at building "causal models"). While the Simon-Blalock version of causal modelling pays lip service to regression coefficients, most applications of causal modelling never deal with regression coefficients; they are devoted to seeing whether certain partial correlations go to zero.[14] And just as simple correlations can mislead, so can partial correlations.

Finally, it is common practice in using data based on geographic units to "deflate" the values of a variable by dividing it by the population of the geographic unit, and then to correlate the deflated variables. For example, a typical ratio or index correlation (as they are often called) is the correlation between welfare expenditures per capita and income per capita, computed over the fifty states. Similarly, most of the correlations computed in various collections of data for the nations of the world are ratio correlations based on "per-capitized" variables. Ratio correlations can be misleading, and the questions they are designed to answer can be more usefully framed as regression problems.[15]

[14] See Hugh Donald Forbes and Edward R. Tufte, "A Note of Caution in Causal Modelling," *American Political Science Review*, LXII (December 1968), 1258-64, and further discussion at 1269-71.

[15] See Wallis and Roberts, 546-56, on the hazards of ratios. The problem was discussed by Karl Pearson, "Mathematical Contributions to the Theory of Evolution—On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs," *Proceedings of the Royal Society of London*, LX (1897), 489-98. See also Edwin Kuh and John R. Meyer, "Correlation and Regression Estimates When the Data are Ratios," *Econometrica*, 23 (October 1955), 400-16; and F.E.A. Briggs, "The Influence of Errors on the Correlation of Ratios," *Econometrica*, 30 (January 1962), 162-77.

Often, then, correlation coefficients are very inadequate tools of analysis. Scatterplots and regression coefficients are more useful. Regression coefficients are certainly preferable in situations where increments along the scales of each variable make some sort of substantive sense.[16] In general, the best procedure is to use both correlation and regression coefficients when they are meaningful. Unfortunately, for all the correlation coefficients reported in *Quantitative International Politics* and in *Politics, Economics, and the Public*, no author reported a single regression coefficient.

V. The Importance of Fitting Lines to Data

Political scientists have tended to be diverted by what are essentially secondary issues in data analysis—significance tests and the distinctions among various levels of measurement. Furthermore, I have argued that correlation coefficients have been overworked and their serious defects ignored in most political data analysis. How, then, can we do a better job of data analysis? How can we get out of summarizing and sanctifying and on to discovering and informing?

The answer to these questions, I think, is rarely to be found in a search for new and exotic techniques—especially if we want to communicate our results to more than a tiny fraction of those involved in the discipline. Almost every new issue of a political science journal introduces a new mathematical or statistical technique. At times it appears that new techniques are introduced not because they fit the substantive problem under investigation, but rather because a new computer program is available to produce some results. There is a real and already visible danger of faddishness in applications of quantitative methods. It is not surprising that the limits of such techniques are often underappreciated by their users. Misguided applications of these techniques, however, not only lend a spurious air of certainty to often false conclusions; such misapplications can also lead to an unwarranted distrust of the methods that seem to have produced the conclusions. Both causal modelling and factor analysis may wind up by being the victims of their own advocates. This may be only wishful thinking, however.

[16] Tukey, 35-66. Hubert M. Blalock, Jr., makes a similar argument in his "Causal Inferences, Closed Populations, and Measures of Association," *American Political Science Review*, LXI (March 1967), 130-36. Blalock's application of the argument to the Miller-Stokes data, however, is a most inappropriate example. For some useful applications and contrasts between standardized and unstandardized regression coefficients, see Hayward R. Alker, Jr. and Bruce Russett, "Multifactor Explanations of Social Change," in Russett and others, *World Handbook and Political and Social Indicators* (New Haven 1964), 311-21.

All in all, political scientists have tended to ignore the useful regression procedures that fit lines to data. For example, none of the authors in *Quantitative International Politics*, a book that is aggressively quantitative, used even the most elementary kind of regression model. In the remainder of this paper, I want to suggest why multiple regression actually is a useful procedure for data analysis.

Fitting lines to relationships between variables (or variables that are transformed) and then examining the deviations off the fitted line by graphs and scatterplots has a good many virtues:

(1) In regression analysis (unlike factor analysis), the research worker must have a fairly specific idea of just what it is he wants to explain. Simply, he has to think in causal terms and know what his response (dependent) and describing (independent) variables are.[17]

(2) Fitting lines to data generates residuals, those parts of the variation in the response variable left unexplained by the describing variables. The effective analysis of residuals is a major tool for discovery.

(3) The resulting regression coefficients, especially if they are unstandardized, occasionally have substantive meaning and policy implications.

(4) Econometricians have built up a large body of useful experience in the application of regression methods to substantive problems. Their advice is sometimes more useful to political scientists than is the advice of other sorts of statisticians.

Finally, let us turn to three important matters in the fitting of lines to data: the analysis of residuals, the transformation of variables, and multicollinearity.

VI. The Analysis of Residuals

Usually the describing variables do not account for all the variation in the response variable. Trying to find something that will explain some of the residual variation helps us discover the unanticipated in a collection of data.

The first step in the analysis of residuals is simply to label them; that is, to calculate the residual for each observation, look at the whole collection of residuals, and see what those observations with residuals of the same size have in common. The next step is to plot the residuals against a wide assortment of things. Here, the computer programmers have let us down a bit, although some regression packages have options

[17] Hayward Alker's "The Long Road to International Relations Theory: Problems of Statistical Nonadditivity," *World Politics*, XVIII (July 1966), at 646-47, has a useful discussion of this point.

for plotting residuals.[18] As far as plotting goes, Tukey and Wilk suggest:

"Kinds of plots of residuals that are very often valuable include (i) plots against fitted, or possible values; (ii) plots against variables which have been employed as a basis for the summarizing fit; (iii) plots against variables, which were not used in the fit, e.g., time; (iv) plots which display residuals identified according to some meaningful characteristic, e.g., according to whether the residual is or is not from an observation which was used in developing the fitted summary; (v) probability plots of ordered residuals, including empirical cumulative distribution plots and plots of empirical quantiles against quantiles of reference distributions, such as the unit normal. While all such plots provide indications of the spread of the body of residuals, it is far more important that they combine palatable summaries of individual residuals with sensitive indications of distributional peculiarities of the entire collection of residuals."[19]

VII. Transformations of Variables

Transformations of variables—that is, any systematic changes in the observed values of the variables made during the course of the analysis —play an important role in the process of fitting lines to data. Transformations are useful because they enable us to use linear techniques to fit rather complicated, nonlinear models to the data, because they often point to substantive results, and, finally, because they help the data satisfy certain statistically desirable properties such as normality and stability of variance.[20]

For example, in the analysis of crossnational aggregate data, the

[18] There are a number of other areas in which current packaged programs are deficient for the needs of social scientists. Two examples here serve to show that we must be careful even though the result came out of the computer. Longley, in a test of commonly used regression programs, found many inaccuracies in the output—including even the wrong sign attached to some coefficients! In this analysis of difficult but real test data (with highly collinear variables), several well-known programs proved accurate to only one or two digits in their estimates of regression coefficients. See James W. Longley, "An Appraisal of Least Squares Programs from the Point of View of the User," *Journal of the American Statistical Association*, 62 (September 1962), 819-41. Second, many cross-tabulation programs have contributed to the frequent misuse of the chi-square test in the analysis of contingency tables. The test is not appropriate for ordered metrics. Of course, it is not entirely the fault of programs when their users dutifully report whatever the printout says.

[19] Tukey and Wilk, 12.

[20] For converting nonlinear models into linear fit problems, see the useful book by N. R. Draper and H. Smith, *Applied Regression Analysis* (New York 1966), chap. 5. The best place to learn about transformations is in the informative and straightforward essay by Joseph B. Kruskal, "Transformations of Data," *International Encyclopedia of the Social Sciences* (New York 1968), vol. 16, 182-93.

logarithm of the variable is often used in place of the actual value of the variable, for convenience and also to meet statistical assumptions. When the variables are logged, moreover, there is an additional gain in terms of the substantive interpretation of the regression analysis. In the two-variable case, with both variables logged, we have

$$\log Y = b \log X + c \,.$$

It can be shown that b, the estimate of the slope, is the elasticity of Y with respect to X.[21] Thus b estimates the proportional change in Y resulting from a proportional change in X. An increase of 1% in X, then, produces under this model a change of b% in Y. This neat interpretation of the slope is an extra benefit of the use of the log transform.

Transformations are also useful in the analysis of percentages. For example, is it always meaningful to assume that the difference between 50% and 60% (an increase of 10%) has the same meaning as the 10% difference between 85% and 95%? The research worker should, when it is appropriate, take into account that a 10% increase starting from 85% as the base is actually often a bigger and more important substantive change than the same amount of percentage-change beginning at an initial level of 50%. If this reasoning is correct, then the tails of percentage distributions should be stretched by transformations of percentages. A number of options are available.[22]

VIII. Multicollinearity

If two or more describing variables are highly intercorrelated, then it is difficult and perhaps impossible to assess their independent effects on the response variable. As the correlation between two independent variables approaches unity, it becomes impossible to tell one variable from the other. The difficulty, called multicollinearity, not only affects the estimates of partial slopes and partial correlations in multiple regression procedures; it also similarly weakens inferences based on cross-tabulations.[23] While occasionally the use of additional information may alleviate the problem, it often happens that when the social scientist must rely on "experiments" performed by nature, he will be unable to

[21] See Alker and Russett, 311-13; also J. Johnston, *Econometric Methods* (New York 1963), 44-52.

[22] See J. B. Kruskal and the references cited there. Another useful discussion is Carl I. Hovland, Arthur A. Lumsdaine, and Fred D. Sheffield, "A Baseline for Measurement of Percentage Change," in Paul Lazarsfeld and Morris Rosenberg, eds., *The Language of Social Research* (Glencoe, Ill. 1955), 77-82.

[23] J. Johnston, 201-07; Hubert M. Blalock, Jr., "Correlated Independent Variables: The Problem of Multicollinearity," *Social Forces*, 62 (December 1963), 233-37; and Donald E. Farrar and Robert R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited," *Review of Economics and Statistics*, 49 (February 1967), 92-107.

obtain the independent variation necessary to assess the independent effects of his explanatory variables.

The problem of multicollinearity has a number of obvious consequences. Some theories that assert the importance of one variable over another, while theoretically testable, are actually incapable of being tested if the describing variables are highly intercorrelated. For example, it may be desirable to separate out the independent effects of economic development and social mobilization on a particular response variable such as military intervention in politics. Yet, for cross-section data in Latin America, the correlation between economic development and social mobilization is 0.89, based on twenty-one observations. Under these circumstances it is simply impossible to assess reliably the independent effects of these two variables.[24] No statistical method—cross-tabulation, regression, path coefficients, or what have you—will break this "multicollinearity deadlock."[25]

An additional danger of multicollinearity is that the analysis, in most cases, can be done as usual. The estimates generated when the describing variables are collinear are, however, subject to large instabilities. The following signs, among others, help alert us to the presence of multicollinearity: (1) high correlations among describing variables; (2) a sizable multiple correlation for the overall regression, but with no particular regression coefficient reaching significance; and (3) large changes in the values of the regression coefficients when new variables are added to the regression.

IX. Conclusion

We can, I think, learn a great deal about politics and society through the statistical analysis of the appropriate data. There are now enough good examples to provide considerable support for this assertion. Techniques of quantitative analysis can help us inform and discover as well as summarize and confirm. We must hope that fewer and fewer applications of quantitative methods will aim only to impress or sanctify.

[24] See Robert D. Putnam, "Toward Explaining Military Intervention in Latin American Politics," *World Politics*, xx (October 1967), 94-95. The finding that economic development is positively correlated with military intervention after the effect of social mobilization is removed is unfortunately not testable because of the high instability of the partial correlation due to multicollinearity. Another example of the problem is discussed in Forbes and Tufte, 1262-64.

[25] Johnston, 207. See Farrar and Glauber for discussion of some modest palliatives.