How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science

Citation	King, Gary. 1986. How not to lie with statistics: Avoiding common mistakes in quantitative political science. American Journal of Political Science 30(3): 666-687.
Published Version	http://www.wiley.com/bw/journal.asp?ref=0092-5853
Accessed	February 18, 2015 9:35:45 PM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:4455012
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

WORKSHOP

How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science[*]

Gary King, *New York University*

This article identifies a set of serious theoretical mistakes appearing with troublingly high frequency throughout the quantitative political science literature. These mistakes are all based on faulty statistical theory or on erroneous statistical analysis. Through algebraic and interpretive proofs, some of the most commonly made mistakes are explicated and illustrated. The theoretical problem underlying each is highlighted, and suggested solutions are provided throughout. It is argued that closer attention to these problems and solutions will result in more reliable quantitative analyses and more useful theoretical contributions.

One of the most glaring problems with much quantitative political science is its uneven sophistication and quality. Mistakes are often made but rarely noticed. In journal submissions, conference presentations, and student papers, problems occur with even more frequency. Having observed this situation for a few years, I noticed several patterns. First, the *same* mistakes are being made or "invented" over and over. Second, to refer a substantively orientated political scientist to an article in *Econometrica, The Journal of the American Statistical Association,* or even *Political Methodology* is to give advice that either is not helpful or is not followed. These problems are more than technical flaws; they often represent important theoretical and conceptual misunderstandings.[1] However, in most cases, there are relatively simple solutions that can reduce or eliminate bias and other statistical problems, improve conceptualization, make the analysis easier to interpret, and make the results more general.

In order to address these concerns, this paper presents proofs and illustrations of some of the most common statistical mistakes in the political science literature, along with theoretical arguments and suggested

[*] An earlier version of this article first appeared at the annual Political Science Methodology Society conference, Berkeley, California, July, 1985. I appreciate the comments from the participants at that meeting, particularly those of Christopher Achen and Nathaniel Beck. Thanks also to my colleagues at New York University, particularly Larry Mead, Bertell Ollman, and Paul Zarowin. Arthur Goldberger, Herbert M. Kritzer, Ann R. McCann, Charles M. Pearson, Lyn Ragsdale, the editors, and the anonymous reviewers were also very helpful.

[1] An example of a minor technical mistake is using ordinal level independent variables with statistics that assume interval level data. I refer to this as "minor" because it usually (although not always) has little substantive consequence and because it does not represent a *conceptual* misunderstanding.

corrections. It specifically omits problems with the newest and fanciest statistical techniques for two reasons. First, the problems considered below form the theoretical and statistical foundation to the more sophisticated methodologies; finding and filling cracks in the foundation should logically and chronologically precede the painting of shingles and shutters. Second, the great variety of newer techniques are being used by relatively few political scientists; thus, any criticism of the new techniques will apply only to a small audience. Although important, I will leave the newer techniques for a future paper.

For each quantitative problem, I describe (1) the mistake, (2) the proof, and (3) the interpretation. The proofs, appearing in footnotes or appendices when excessively technical, are formal versions of, as well as algebraic or numerical evidence for, the assertions made in my discussion of the mistake. Emphasis here is on the intuitive, so generality is often sacrificed in order to improve conceptual understanding. The final section includes a brief summary and gives implications of mistakes in the context of proposed solutions. Some sections are too brief to be divided into this triad and are therefore combined. This sort of methodological retrospective has been done in other disciplines, but although we can learn from some of these, most do not address problems specific enough to political science research. (See, for example, Leamer, 1983a; Smith, 1983; Friedman and Phillips, 1981; and Hendry, 1980; Gurel, 1968).[2]

Over three decades ago, Darrell Huff (1954) explained, in a book by the same name, *How to Lie With Statistics.* Because of the systematic precision required, we should realize by now that it is a lot harder (knowingly or not) to lie (and get away with it) with statistics than without them.

Regression on Residuals

The Mistake. Suppose that y were regressed on two sets of independent variables X_1 and X_2.[3] The coefficients to be estimated are in the parameter vectors β_1 and β_2 in model 1:

$$E(y \mid X_1, X_2) = X_1\beta_1 + X_2\beta_2 \tag{1}$$

The standard and appropriate way to estimate β_1 and β_2 in model 1 is by running a multiple regression of y on X_1 and X_2. The result

$$y = X_1 b_1 + X_2 b_2 + e \tag{2}$$

[2] I do not cite every methodologically flawed political science work in this paper because the purpose here is to improve future research and to facilitate critical reading of all research. There is little gained by berating those on whose research we are trying to build.

[3] The word "regressed" is sometimes misused. Reading a regression equation from left to right, we say, "the dependent variable is regressed on the independent variables." In the text, y is the dependent variable; X_1 and X_2 each represent a set of several independent variables.

is the least squares (LS) estimator. The sample estimates in equation 2 are used to infer to the population parameters in equation 1.

Now consider an (incorrect) alternative procedure, called here the regression on residuals (ROR) estimator. This is a method of estimating β_1 and β_2 often "invented" by first regressing y on X_1, resulting in this equation:

$$y = X_1 b_1^* + e_1 \tag{3}$$

where b_1^* is the first ROR estimator.

We then regress e_1, the residuals, on the second set of explanatory variables, X_2, yielding

$$e_1 = X_2 b_2^* + e_2 \tag{4}$$

where b_2^* is the second ROR estimator.

The mistaken belief is that b_2^* from the second regression in equation 4 is equal to b_2 from equation 2; that is, since we have "controlled" for X_1 in equation 3, the result is the same as if we had originally computed 2. As is demonstrated in the proof appearing in appendix A, this is not true. The ROR estimator b_1^* in equation 3 is a biased estimate of β_1, since the equation does not control for X_2. This is the well-known omitted variables bias.[4] Since e_1—the residuals from equation 3 and the dependent variable in equation 4—is calculated from the biased ROR estimator b_1^*, it too is biased. Thus, it follows that b_2^* is also biased, since it is calculated from the regression of the biased e_1 on the second set of explanatory variables X_2.[5]

The Interpretation. Except for two very special cases, the ROR estimator is not the same as the ordinary least squares estimator and by itself has no useful interpretation. b^* is also a biased estimate of β in model 1. In order to estimate β_1 and β_2 correctly in model 1, both sets of variables X_1 and X_2 should be put in the regression simultaneously. This gives an estimate (b_1) of the influence of X_1 on y (controlling for X_2), and an estimate (b_2) of X_2 on y (controlling for X_1).

An implication of this result is that one should not make too much of any interpretation of the residuals from a regression analysis. If it appears from an analysis of the residuals that some variable X_3 is missing, then X_3 may be missing, but it is not possible to draw fair conclusions about the

[4] The bias does not occur when either $\beta_2 = 0$ or X_1 and X_2 are uncorrelated or both.

[5] Sometimes this process is continued: The second set of residuals e_2 is regressed on another set of explanatory variables, X_3, producing another ROR estimator and another set of residuals. This process has been extended to many stages, but I only consider the first two in the text. In the multi-stage ROR estimator, the bias is confounded even further.

influence of X_3 on y unless X_3 were actually measured and the full equation were estimated.

An example is Achen's (1979) result that "Normal Vote" calculations are inconsistent: the Normal Vote was determined by a two-step process, roughly analogous to using the ROR estimator.

In the statistical literature, the ROR estimation procedure is called "stepwise least squares." However, "stepwise regression" is very different from this procedure—although it is no less problematic.[6]

The Race of the Variables

In this section, the use of standardized coefficients ("beta weights"), correlation coefficients (Pearson's correlation), and R^2 ("the coefficient of determination") are challenged. In most practical political science situations, it makes little sense to use these statistics. They do not measure what they appear to; they substitute statistical jargon for political meaning; they can be highly misleading; and in nearly all situations, there are better ways to proceed.

The Race (1): Standardized Fruit

The Mistake: Apples, Oranges, and Perceptions. Imagine a situation where a researcher wanted to explain y, the number of visits to the doctor per year. The explanatory variables were X_1, the number of apples eaten per week, and X_2, the number of oranges eaten per week. The multiple regression equation was then estimated to be:

$$\hat{y} = 10 - 1.5X_1 - 0.25X_2. \qquad (5)$$

[6] Stepwise regression (which has been called "unwise regression" [Leamer, 1985] or might be called a "Minimum Logic Estimator"), allows computer algorithms to replace logical decision processes in selecting variables for a regression analysis. There is nothing wrong with fitting many versions of the same model to analyze for sensitivity. After all, the goal of learning from data is as noble as the goal of using data to confirm *a priori* hypotheses. However, some *a priori* knowledge, or at least some logic, always exists to make selections better than an atheoretical computer algorithm. Edward Leamer (1983b, p. 320) has noted, "Economists have avoided stepwise methods because they do not think nature is pleasant enough to guarantee orthogonal explanatory variables, and they realize that, if the true model does not have such a favorable design, then omitting correlated variables can have an obvious and disastrous effect on the estimates of the parameters." At the very least, stepwise regression, even if occasionally useful for special purposes, need not be presented in published work (see Lewis-Beck, 1978). The use of stepwise regression has caused an additional curious mistake. It is often said that the order in which variables are entered into a regression equation influences the values of the coefficients. A cursory look at the equations used in the estimation (or at a sample computer run) will show that this is wrong. What does change is dependent upon the order variables are entered is the marginal increase in the R^2 statistic.

For every additional apple one eats per week, the average number of visits to the doctor per year decreases by one and a half. For each additional orange one eats, they decrease by one quarter of a visit.

This hypothetical researcher now would like to make a statement about the comparative worth of apples and oranges in reducing doctor visits. He then asks the resident political science methodologist whether she can help him compare apples and oranges. The methodologist says that the answer depends upon the researcher stating his question more precisely. If the researcher means: "I have only enough money for one apple or one orange, and I want to know which will make me healthier," then the answer is probably the apple. But suppose an apple costs 50 cents, while an orange costs only five cents. In this case, the researcher might ask, "What is the best use of my last dollar?" Here the decision would have to be in favor of the orange: For one dollar spent on two apples, doctor visits would decrease by about three, whereas the same dollar spent on 20 oranges would decrease doctor visits by five on average.

Assuming the question is stated precisely enough, these comparisons make some sense. But they make sense only because there is a common unit of measurement—a piece of fruit or an amount of money. Suppose then that the researcher told the methodologist that he had torn off the computer printout just prior to the last coefficient estimate. The real equation, he explained, includes X_3, the respondent's perception of doctors as beneficial, measured on a scale ranging from 1 (not beneficial) to 10 (very beneficial). The estimated equation should have appeared as this:

$$\hat{y} = 10 - 1.5X_1 - 0.25X_2 + 2X_3 \tag{6}$$

The researcher now asks whether this means that perceptions are "more important" than apples. After all, he says, 2 is greater than 1.5. Any methodologist worth her 8087 chip would object to this, she asserts. In fact, were one to take this comparison to its logical extreme, one would conclude that perceiving doctors as more detrimental is more health-producing than eating an apple. Although both regressors seek to explain the same dependent variable, they are neither measured on, nor can they be converted to, meaningfully common units of measurement.

This is precisely the point: *Only when explanatory variables are on meaningfully common units of measurement is there a chance of comparison.* If there is no common unit of measurement, there is no chance of meaningful comparison.

However, there is another sense in which even "common-unit" comparisons are unfair. The apple coefficient, for example, represents the effect of apples (holding constant the influence of oranges and perceptions). The estimated coefficient for oranges has a different set of control variables (since it includes apples and not oranges). This may make a comparison between apples and oranges more difficult, if not logically impossible.

The Mistake Continued: Standardized Fruit. Convinced about not comparing unstandardized coefficients, our hypothetical researcher proposes using the standardized coefficients on his computer printout. The methodologist retorts that, if it is of little use to compare apples and perceptions, then it is of considerably less use to compare standardized apples and standardized perceptions. Standardization does not add information. If there were no basis for comparison prior to standardization, then there is no basis for comparison after standardization.

A relatively common rebuttal is that for explanatory variables with unclear or difficult-to-understand units of measurement, standardized coefficients should increase interpretability. The problem is that if the original data were meaningless, then the standardized regression coefficients are precisely as meaningless; if standardized coefficients do not add information, they certainly do not add meaning. "To replace the unmeasurable by the unmeaningful is not progress" (Achen, 1977, p. 806).

Using a superscript "*s*" to denote standardized variables, I present the results for our hypothetical case:[7]

$$\hat{y}^s = -0.9X_1^s - 0.2X_2^s + 0.5X_3^s \qquad (7)$$

We now must interpret equation 7 to mean, for example, that as we eat one additional standard deviation of apples, the number of visits to the doctor decreases by nine tenths of a standard deviation—not a very appealing conceptualization.

Three observations: First, standardizing makes the coefficients substantially more difficult to interpret. Second, standardization still does not enable us to compare this first effect to the one-half standard deviation increase in doctor visits resulting from a one standard deviation increase in perceptions of doctors.

Third, and most serious, while the original coefficients are estimates of the relationships between the respective explanatory variables and the dependent variable (controlling for the other explanatory variables), the standardized variables are measures of this relationship as well as of the variance of the independent variable. Since researchers are typically interested in measuring only the relationship, or at least interested in the two separately,

[7] There are two methods that can produce the same standardized coefficients: (1) standardize each of the original variables (subtract the sample mean and divide by the sample standard deviation) and run a regression on these standardized variables; or (2) run a regression and multiply each unstandardized coefficient by the ratio of the standard deviation of the respective independent variable to the standard deviation of the dependent variable.

it makes little sense to use standardized variables. A simple numerical proof will demonstrate this point.[8]

The Proof. Imagine a simple experiment where only three observations on one dependent and one independent variable are taken. The observations are $y' = \{5, 5, 6\}$ and $X' = \{2, 4, 4\}$. Calculated from these three observations with a constant term included, the regression is:

$$\hat{y} = 4.5 + 0.25X \tag{8}$$

and the standardized coefficients are:

$$\hat{y}^s = 0.50X^s \tag{9}$$

Suppose further that another year went by, and another data point was collected on y {9.5} and on X {20}. Because this random draw worked out well, the unstandardized coefficients in equation 8 do not change at all with the introduction of this additional observation. However, the new observation increases the sample standard deviation of X from 1.16 to 8.39 (which is what one would generally expect as n increases). Although this did not change the original coefficients in equation 8, the standardized coefficient nearly doubles in the four observation case (compare equations 9 and 10):

$$\hat{y}^s = 0.97X^s \tag{10}$$

Under situations with different variances of the independent variables but identical relationships, the standardized coefficient is constrained only to have the same sign as the unstandardized coefficient. Standardized coefficients may be either under- or over-estimates. This intuitive proof extends directly to situations with multiple independent variables.

The Interpretation. In summary, standardized coefficients are in general (1) more difficult to interpret, (2) do not add any information that may help to compare effects from different explanatory variables, and (3) may add seriously misleading information. The original, unstandardized coefficients are meaningful and are not subject to these problems, although they generally cannot be compared for importance.

There are two important qualifications to these points. First, if one must include a variable that is difficult to interpret as a control, then perhaps standardizing *just this variable* would capitalize on the standardized coefficient's simpler descriptive properties (Blalock, 1967a). This partial standardization procedure is certainly better than standardizing all

[8] Kim and Mueller (1976) also show that changes in the covariances of the included variables and of the variances of the included and excluded variables in a system of equations also affect the standardized (but not the unstandardized) coefficients.

the variables.[9] Second, some argue that standardized measures seem to be the more natural scale for variables like test scores.[10] For example, Hargens (1976) argues that standardized coefficients can sometimes be structural parameters. Although Kim and Ferree (1981) successfully refute most of this argument on theoretical grounds, there is one sense in which it may be correct for some studies. To make this point, it is useful to consider a very different type of standardization commonly used and generally accepted in economic studies of time series data.

The raw consumer price index (CPI_t) is not usually included in regression models for two reasons. First, the series is nonstationary and may, therefore, lead to spurious findings. Second, for example, an increase in the price of a typical market basket of food from \$10.00 to \$11.00 is likely to have more of an influence on any dependent variable than if the increase were from \$100.00 to \$101.00. For both of these reasons, the proportional change in CPI_t is used; this "standardized" measure is commonly called the inflation rate.[11] In this case, the standardized variable is usually considered more natural and substantively meaningful than the "unstandardized" CPI_t.

In a similar manner, subtracting the sample mean from a variable under analysis and dividing it by the sample standard deviation may be the more natural measure for some concepts, particularly for some psychological scales and attitudinal measures. In part, it may even be a matter of personal taste and custom (Blalock, 1967b). However, decisions about whether each variable is to be standardized should be made and justified on an individual basis rather than "a habitual reliance on the standardized coefficients" (Kim and Ferree, 1981, p. 207). Just as we should not routinely calculate proportional changes for every variable in a time series analysis, variables in cross-sectional analyses should not be automatically standardized.

A more important and final point is that most times scholars are not interested in finding out which variable will win the race. Most often it is theoretically "good enough" to say that even after controlling for a set of

[9] If the *dependent* variable is too difficult to understand, then I would give up on the regression, collect better data, or try to figure out a more meaningful interpretation.

[10] As an example of the problem this sometimes causes, consider the Educational Testing Service's (ETS's) standardized Graduate Record Examination (GRE). University admission offices across the country make important decisions based in part on small differences in scores on this exam, whereas ETS reports that the GRE can only correctly distinguish students who are more than one hundred points apart (on a scale from 200 to 800) two out of three times (i.e., a 66% confidence interval!). Perhaps if this score were not standardized or if there were a more meaningful substantive interpretation, we would be better prepared to use GREs for admission decisions.

[11] The most intuitive way to calculate the inflation rate is as $(\mathrm{CPI}_t - \mathrm{CPI}_{t-1})/\mathrm{CPI}_{t-1}$, but a nearly exact measure, which for technical reasons is actually better and is used most everywhere, is $\log(\mathrm{CPI}_t) - \log(\mathrm{CPI}_{t-1})$. See King and Benjamin (1985) for a political application.

variables (i.e., plausible rival hypotheses, possible confounding influences), the variable in which we are interested still seems to have an important influence on the dependent variable. This is precisely the empirical evidence for which we search to substantiate or refute our theoretical expectations. Usually, little political understanding is gained by hypothesizing a winner in a race of the variables.

The Race (2): The Correlation Problem

The Mistake. Many great things are attributed to the simple correlation coefficient. It purportedly needs to assume covariation, while regression must assume causation. The specific statistical assumptions are thought to be less severe than for regression. It is said to be a better guide when one's theory argues only that "the variables generally go together" rather than there being a "one-to-one, cause and effect relationship." It also supposedly makes results easier to interpret.

Each of these statements is false. There are several approaches to describing why these common arguments are invalid (Tufte, 1974). Two are most useful for present purposes.

The Proof. Consider, first, the case of a standardized coefficient on one independent and one dependent variable. Through some simple algebraic manipulation, it can be shown that this standardized coefficient is equal to the correlation coefficient.[12] Thus, *every argument that applies to the standardized regression coefficient, applies also to the correlation coefficient.*

Next, consider the population parameters to which the sample correlation attempts to infer. The most likely relevant probability distribution is the bivariate normal, which has five parameters: the marginal mean and variance for each variable and ρ, the population correlation coefficient. The problem is that if r were considered an estimator of ρ we would need to assume that x and y were drawn from a bivariate normal distribution. Since the marginal distributions of a bivariate normal distribution are normally distributed, we would need to make all the assumptions of re-

[12] With no loss of generality, assume each variable has a mean of zero. This proof demonstrates that the standardized coefficient (b^s) for one independent and one dependent variable is equal to the correlation coefficient (r):

$$b^s = b\frac{S_x}{S_y} = (x'x)^{-1}x'y\frac{\sqrt{\Sigma x_i^2}}{\sqrt{\Sigma y_i^2}} = \frac{\Sigma \mathrm{x}_i \mathrm{y}_i}{\Sigma x_i^2} \cdot \frac{\sqrt{\Sigma x_i^2}}{\sqrt{\Sigma y_i^2}}$$

$$= \frac{\Sigma x_i y_i}{\sqrt{\Sigma x_i^2}\sqrt{\Sigma y_i^2}} = r$$

For the case of multiple independent variables, standardized coefficients are not the same as correlation coefficients or partial correlation coefficients. The results presented here therefore are not completely general, but the substantive conclusions and recommendations still apply.

gression and, in addition, the assumptions that X is normally distributed and that x and y are jointly normally distributed. In many political science examples, this is unreasonable. For example, any use of a dichotomous independent variable (male/female, agree/disagree, etc.) violates the assumption. Moreover, one can use regression, make fewer assumptions, and get more reasonable and interpretable results.

The Interpretation. All of the problems attributed to standardized coefficients apply to correlation coefficients.

Furthermore, there is nothing in statistical theory that attributes causal assumptions to regression coefficients; regression is simply a sample estimate of a (population) conditional expected value. The assumptions are about the conditional probability distribution, not about the influence of x on y. Nothing can or should stop an applied researcher from stating that x causes y, but it is crucial to understand that statistical analysis does not usually provide evidence with which to evaluate this assertion (see Granger (1969) and Sims (1980) for more direct attempts).

There is also nothing that attributes causal assumptions to the correlation coefficient. Correlations are sample estimates of the population parameter ρ from the bivariate normal distribution. Thus, arguments about causality, association, and correlation are not required for either regression or correlation and do not form a basis for choosing between the two.

Furthermore, as a result of the distributional requirements, the assumptions for correlation coefficients are far more demanding than for regression analysis. Unstandardized regression coefficients are almost always the best option.

The Race (3): Coefficient of Determination?

R^2 is often called the "coefficient of determination." The result (or cause) of this unfortunate terminology is that the R^2 statistic is sometimes interpreted as a measure of the influence of X on y. Others consider it to be a measure of the fit between the statistical model and the true model. A high R^2 is considered to be proof that the correct model has been specified or that the theory being tested is correct. A higher R^2 in one model is taken to mean that that model is better.

All these interpretations are wrong. R^2 is a measure of the spread of points around a regression line, and it is a poor measure of even that (Achen, 1982). Taking all variables as deviations from their means, R^2 can be defined as the sum of all $\hat{y}^2$ (the sum of squares due to the regression) divided by the sum of all y^2 (the sum of squares total):

$$R^2 = \frac{\hat{y}'\hat{y}}{y'y} = \frac{b'X'Xb}{y'y} = \frac{\Sigma Xy}{\Sigma X^2 \Sigma y^2}$$

where the last equation moves from general notation to that for one independent variable.

Note, however, that this is precisely the square of the correlation coefficient (or the square of the standardized regression coefficient given in footnote 12). Therefore *all of the criticisms of the correlation and standardized regression coefficients apply equally to the R^2 statistic.*

Worse, however, is that there is no statistical theory behind the R^2 statistic. Thus, R^2 is not an estimator because there exists no relevant population parameter. All calculated values of R^2 refer only to the particular sample from which they come. This is clear from the standardized coefficient example in preceding paragraphs, but it is more graphically demonstrated in two (x, y) plots by Achen (1977, 808). In the first plot $R^2 = 0.2$. In the second plot, the fit around the regression line is the same, but the variance of X is larger; here $R^2 = 0.5$.

Ad hoc arguments for R^2 are often made in the form of the researcher's questions and the methodologist's answers:

Q: How can I tell how strongly my independent variables influence my dependent variable without R^2?

A: Interpret your unstandardized regression coefficients.

Q: But how can I tell how good these coefficients are?

A: The standard errors are estimates of the variance of your estimates across samples. If they are small relative to your coefficients, then you should be more confident that similar results would have emerged even if a sample of 1500 different people were interviewed.

Q: But how can I tell how good the regression is as a whole?

A: If you want to test the hypothesis that all your coefficients are zero, use the F-test. More complex hypotheses about different theoretically relevant linear combinations of coefficients (e.g., that the first three coefficients are jointly zero, or that the next two add to 1.0) can also be tested. R^2 is associated with, but is a poor substitute for, test statistics.

Q: O.K. I guess I really mean to ask: How can I assess the spread of the points around my regression line?

A: There is nothing intrinsically or politically interesting in the spread of points around a regression line. If you are interested in the precision with which you can confidently make inferences, then look at your standard errors. Alternatively, you might be interested in the precision of within-sample and out-of-sample forecasts. Forecasts correspond to the regression line (or to the extrapolated line for out-of-sample forecasts), given specified values of your explanatory variables. It is perfectly reasonable to estimate and then make probabilistic statements about the forecasts or even to calculate forecast confidence intervals. Surely if the observed point spread is large, the confidence interval will also be large. However, R^2 is also a poor substitute for going directly to confidence intervals.

Q: But do you really want me to stop using R^2? After all, my R^2 is higher than that of all my friends and higher than those in all the articles in the last issue of the *APSR!*

A: If your goal is to get a big R^2, then your goal is not the same as that for which regression analysis was designed. The purpose of regression analysis and all of parametric statistical analyses is to estimate interesting population parameters (regression coefficients in this case). *The best regression model usually has an* R^2 *that is lower than could be obtained otherwise.*

If the goal is just to get a big R^2, then even though that is unlikely to be relevant to any political science research question, here is some "advice": Include independent variables that are very similar to the dependent variable. The "best" choice is the dependent variable; your R^2 will be 1.0. Lagged values of y usually do quite well. In fact, the more right-hand-side variables included the bigger your R^2 will get.[13] Another choice is to add variables or selectively add or delete observations in order to increase the variance of the independent variables.

These strategies will increase your R^2, but they will add nothing to your analysis, nothing to your understanding of political phenomena, and nothing useful in explaining your results to others. The general strategy of analysis will likely destroy most of the desirable properties of regression analysis.

Q: Is there anything useful about R^2?

A: Yes. There is at least one direct use and several indirect uses of R^2. You can directly apply and evaluate R^2 when comparing two equations with different explanatory variables and identical dependent variables. The measure is, in this case, a convenient goodness-of-fit statistic, providing a rough way to assess model specification and sensitivity. For any one equation, R^2 can be considered a measure of the proportional reduction in error from the null model (with no explanatory variables) to current model. As such, it is a measure of

[13] It is possible, but unlikely, for the R^2 to stay the same; in any case, it will never decrease as more variables are added. More generally, as the number of variables approaches the number of observations, R^2 approaches 1.0.

> the "proportion of variance explained," and, although this interpretation is commonly used, it is not clear how this interpretation adds meaning to political analyses.

There are also a variety of indirect "uses" for R^2. It is often true that a high R^2 is accompanied by small standard errors, large coefficients, and narrow confidence intervals. Thus, a higher R^2 is generally good news; this is the reason why, *ceteris paribus,* R^2 does not always mislead. However, most of the useful information in R^2 is already available in other commonly reported statistics. Furthermore, these other statistics are more accurate measures: They can directly answer theoretically interesting questions. R^2 cannot. Of course, when one reads someone else's work, R^2 may be a useful interpretive substitute if some of the more accurate measures were not calculated. Consequently, although the odds of being misled are substantially higher with R^2 than with these other statistics, it is just as well that R^2 is routinely reported. It is the use of this information that should be changed.

Confusion with Dichotomous Variables

In this section, I discuss common misuses of dichotomous variables. First, I consider the relationship between analysis of variance and regression in handling dichotomous independent variables. Then, I present common mistakes in using dichotomous dependent variables. Finally, I attempt to alleviate confusion about using dichotomous variables and mistaking dependent variables for independent variables in factor analysis.

(1) Dichotomous Independent Variables

The Mistake. Consider a case where there are two populations with means μ_1 and μ_2, from which random samples sizes n_1 and n_2 are taken. The populations could be male and female, agree and disagree, Republican and Democrat or anything that could be represented by a meaningful explanatory dichotomous variable. A common problem is to test the hypothesis that the means are equal ($\mu_1 - \mu_2 = 0$). In this case, the first thing we do is calculate the means, $\bar{y}_1$ and $\bar{y}_2$, of the two samples.

There are three approaches to this problem: (1) a difference in means test, (2) an analysis of variance (ANOVA) model, and (3) a regression model. Justifications for choosing one of these models over the others are often given. The difference in means test is sometimes seen as a quick way to get a feel for the data. ANOVA and the difference in means have been credited with requiring less restrictive assumptions about the data. Some think ANOVA can be safely used with dichotomous dependent variables; others say that ANOVA and not regression allows dichotomous independent variables.

These assertions are false. In fact, the three techniques are intimately related—conceptually, statistically, and even algebraically. The simplest but least general of the three is the difference in means test. Let y be a vector of observations from both populations and X be an indicator variable. Let the value for the first population be -1 and the value for the second be 1. (These values are arbitrary choices that make later computation easier.) Then the model is

$$E(y \mid X = -1) = \mu_1$$

$$E(y \mid \mathrm{X} = 1) = \mu_2 \qquad (11)$$

The obvious sample statistic is the difference in the sample means, which, after dividing by the standard error of this difference, follows a t-distribution.[14]

Analysis of variance (ANOVA) is a somewhat more general way to deal with this problem. The theoretical model is $E(y) = \mu + \delta_i$, where μ is the grand mean of both populations, $i = 1, \ldots, G$, where G is the number of populations, and δ_i is the deviation from the grand mean for population i. We impose the restriction that $\sum_{i=1}^{G} \delta_i = 0$. In the special case of $G = 2$, $\delta_1 = -\delta_2$. The model can be restated for each population as

$$E(y \mid X = -1) = \mu + \delta_1$$

$$E(y \mid X = 1) = \mu + \delta_2 = \mu - \delta_1 \qquad (12)$$

The sample estimate of μ is $\bar{\mathrm{y}}$ and of δ_i is d_i. By definition,

$$\bar{y} + d_1 = \bar{y}_1 \text{ and } \bar{y} + d_2 = \bar{y}_2.$$

These means are, of course, identical to those that estimate model 11, but d_1 and d_2 represent deviations from the sample "grand" mean. The representation is slightly different, but the interpretation should be exactly the same. The test statistic for the hypothesis that $\delta_1 = \delta_2 = 0$ follows the F-distribution, which is a trivial generalization of the t-distribution used for the difference in means test.[15]

The final and most general approach to this problem is with regression analysis (the general linear model). The model of interest here is $E(y \mid X) = \beta_0 + \beta_1 X$, with X taking on the value -1 for the first population and 1 for the second. The model is defined, for each population, as:

[14] The choice for an estimate of the standard error depends upon whether the two samples are independent. Although I consider here only the case of independence, there are straight-forward generalizations to the case of "nonspherical" disturbances.

[15] Squaring a variable with a t-distribution ($df = k$) yields a variable with an F-distribution ($df_1 = k$, $df_2 = 1$).

$$E(Y|X=-1)=\beta_0-\beta_1$$
$$E(Y|X=1)=\beta_0+\beta_1 \tag{13}$$

The sample estimates of β_0 and β_1 are b_0 and b_1, respectively. Appendix B proves that b_0 and b_1 are close algebraic relatives of the grand mean ($\bar{y}$) and the deviations from that mean (d_i) from the ANOVA model. Two points are demonstrated in Appendix B. First, b_0 is shown not to be the grand mean except when $n_1=n_2$. Second, b_1 is proven not to be the deviation from the grand mean (d_i), except when $n_1=n_2$.

This inequality between ANOVA and regression only denotes different ways of representing the same underlying relationships. There are no differences in assumptions or empirical interpretation. Note that in the regression model, deviations from the grand mean can be represented in terms of the parameter estimates:

$$\begin{aligned}\bar{y}-\bar{y}_1 &= b_0+b_1[(n_2-n_1)/n]-(b_0+b_1)\\ &= \left[\frac{n_2-n_1}{n}\right]-b_1\end{aligned}$$

$$\begin{aligned}\bar{y}-\bar{y}_2 &= b_0+b_1[(n_2-n_1)/n]-(b_0-b_1)\\ &= \left[\frac{n_2-n_1}{n}\right]+b_1\end{aligned}$$

Thus, for the special case of $n_1=n_2$, $\bar{y}-\bar{y}_1=-b_1$ and $\bar{y}-\bar{y}_2=b_1$ just as in ANOVA. When $n_1\neq n_2$, we interpret $2b_1$ to be estimated difference between the two population means. In fact, $2b_1$ is exactly the parameter estimate for the difference in means test.

Note that in none of these models should dichotomous independent variables be standardized. The consequence of such a calculation is to make the standardized coefficient dependent not only upon the variance of the independent variable (as is always the case) but also upon its mean, since the variance of dichotomy is a function of its mean.

The Interpretation. ANOVA, regression, and the difference of means test are all special cases of the general linear model. The assumptions required of one are required of the others as well. If there are dichotomous dependent variables, none of the techniques are appropriate. If there is a dichotomous independent variable, any one of the three will do. If, as is usually the case with political data, there are both discrete and continuous explanatory variables, then only regression will accommodate the research problem.

There are generalizations of ANOVA that accommodate mixed models like "analysis of covariance," (which is not to be confused with "analysis of covariance structures"). Since, for experimental researchers, ANOVA often seems a more conceptually appropriate model, and since the same data requirements and resulting information is essentially equivalent to regression analysis, the choice between the two is mostly a matter of personal taste.

My view is that for most political science research, regression is a substantially more general model: It incorporates many types of ANOVA in one statistical model (and algebraic formula). Although specification issues apply to all three methods, they are usually only considered in a regression context. In addition, regression is also substantially easier to generalize in order to correct for nonspherical disturbances and other common problems. By comparison, more general ANOVA models can get quite messy when they exist. For this reason, many ANOVA computer programs actually do regression analyses and then transform the results into the ANOVA parameters for presentation.

The point is that for the standard analysis, all three models come from the same general form. Each model provides a different representation of exactly the same information, and correct specification is required of all three. When the analysis is more complicated, the regression model may prove more tractable.

(2) Dichotomous Dependent Variables

The Mistake. The mistake here is using dichotomous dependent variables in regression, ANOVA, or any other linear model. Doing this can yield predicted probabilities greater than one or less than zero, heteroskedasticity, inefficient estimates, biased standard errors, and useless test statistics. Of more importance is that a linear model applied to these data is of the wrong functional form; in other words, it is conceptually incorrect.

Consider, for example, the influence of family income on the probability of a child attending college (measured as a dichotomous, college/no college realization). Hypothesizing a linear relationship in this situation implies that an additional thousand dollars of family income will increase the probability of going to college by the same amount regardless of the level of income. Surely this is not plausible. Imagine how little difference an additional $1000 would make for a family with $1,000,000, or for one with only $500, in annual family income. However, for a family at the threshold of having enough money to send a child to college, an additional thousand dollars would increase the probability of college attendance by a substantial amount. The relationship this implies is a steep regression line (representing a strong effect) at the middle range of income and a relatively flatter line (a weaker relationship) at the extremes. Extending this for all values of income produces the familiar logit or probit S-curve (for an application, see King, in press, 1986).

The solution is to model this relationship with a logit or probit (or some other appropriate non-linear) model. Scholarly footnotes to the contrary, it is not possible to do logit and regression analyses and have them "come out the same." What exactly is meant by "come out the same"? It would be meaningless to compare logit and LS coefficients, standard errors, or test statistics. There is no such thing as R^2 in logit analysis, and although there are analogous statistics, comparisons make little sense; in any case, logit analysis will always have a fit to the data as good as or better than that of LS estimation. The interpretation cannot be the same, since the underlying theoretical models are very different.

There is, however, one proper comparison between LS and logit estimation—between the fitted values of the two models expressed as proportions.[16] A short-hand way to accomplish this for the logit model is by observing the first derivatives of the logit function, $bp(1-p)$, where b is the logit coefficient and p is the initial probability. The problem is that unlike LS, the effect on y for an additional unit increase in X is not constant over the range of X values. This "variable effect" is represented as a nonlinear logit function.[17]

(3) Confusing Dichotomous Independent with Dichotomous Dependent Variables

In the factor analysis model, there are many observed variables from which the goal is to derive underlying (unobserved) factors. A common mistake is to view the observed variables as causing the factor. This is incorrect. The correct model has observable *dependent* variables as functions of the underlying and unobservable factors. For example, if a set of opinion questions asked of the political elite is factor analyzed, underlying ideological dimensions are likely to result. It is the fundamental ideologies that cause the observed opinions, and it is precisely because these ideologies are unobservable that we measure only the consequences of these ideologies.

This has two practical consequences for the researcher. First, variables

[16] When the underlying probability for each observation remains within the 0.25 to 0.75 probability interval, the logit and LS models produce very similar predicted values. However, standard errors and test statistics have little meaning; although they have somewhat more meaning when probabilities are within the 0.25 to 0.75 interval. Of course, projections of the underlying theoretical model are always implausible with LS.

[17] For the special case of only nominal level independent variables, Kritzer (1978a; see also 1978b) shows that a minimum chi-square estimation procedure becomes a very intuitive weighted least squares on tabular data. This article is also a good example of the point made in the previous section—that many different statistical models can be organized under the regression framework.

like race, gender, and age should never be observed variables in a political scientist's factor analysis. It is doubtful that a researcher will ever find that party identification or ideology influences a person's gender or race. Second, since most factor analysis models are linear, they can no more handle dichotomous dependent (observed) variables than can regression analysis models. However, there are nonlinear factor analysis models, which are generalizations of the binary logit model, that may be appropriate in this situation (Christoffersson, 1975).

Reporting Replicable Results

I focus in this section on reporting results of statistical analyses. An erroneous reporting method, if not the most grievous offense, is certainly the most frustrating. After all, if a mistake is made and reported, then it is sometimes possible to assess the damage. If minimum reporting standards are not followed, then the only conclusions that can be drawn are based on blind faith in or rejection of the author's interpretative conclusions and methodological skills. Tabular information conveys information that usually is not (and usually should not be) presented in the text. If the tables are not complete, then the report may be rendered useless.

I have concentrated in this paper primarily on regression analysis, the most frequently used explicit statistical model in political science research and the most frequently abused. As an example, therefore, consider reporting the results of a LS analysis. The required results should be (1) data descriptions (including the unit of measurement for each variable, the unit of analysis, and the number of observations and variables), (2) parameter estimates (regression coefficients and the estimated variance of the disturbances), and (3) the standard errors (measures of the precision of the coefficient estimates). For time-series analyses and certain types of cross-sectional analyses, tests of or searches for nonspherical disturbances (e.g., autocorrelation and heteroskedasticity) should also appear.[18] If joint hypothesis tests are relevant, but not executed, relevant parts of the variance-covariance matrix of the regression coefficients (on the diagonal of which are the squares of the standard errors) should be included. Since they can be derived from the information presented, t-tests, F-tests, goodness-of-fit statistics, and marginal probability levels are optional.

One relatively common violation of these reporting rules is to replace any coefficient not meeting some significance level with "N.S." (Sometimes

[18] Automatic use of the Durbin-Watson statistic in time-series data is better than nothing, but it is far from the best approach. A better procedure is to analyze the autocorrelation and partial autocorrelation functions of the residuals. Although full reporting of these would be excessive, a sentence or two summarizing any odd results would be very helpful.

even the level at which these coefficients are not significant is not reported!) This procedure can be very misleading. In fact, I know of no political science research in which it makes sense to use a precise critical value. Any coefficient that is significant at the 0.05 level is as useful in this discipline as if it were 0.06 or 0.04. To delete and refuse to interpret a coefficient which is 0.01 or 0.001 above a significance level makes little sense. Even if the author has a reason for it, at least readers could be permitted to come to their own conclusions. My recommendation is to present the marginal probability level (the exact "level of significance") for each coefficient, regardless of what it is; the author can argue whatever he or she wants and readers would still be able to draw their own conclusions. Statistical significance and substantive importance have no necessary relationship.

There are many other examples of incomplete tables, misleading footnotes and useless appendices. The best general way to judge the adequacy of reporting is to determine if the analysis can be replicated. It, of course, need not *be* replicated, but in order to contribute methodological and theoretical information to its readers, a paper must report enough information so that the results it gives could be replicated if someone actually tried.

Remarks

This paper reviews some of the more common conceptual statistical mistakes in quantitative political science research. Although many mistakes are caught by perceptive colleagues, many more slip by. Those presented here are among the most systematically problematic. Too often, we *learn* each others' mistakes rather than *learning from* each others' mistakes. Fortunately, in each case, there are plausible reasons for the initial mistaken "invention" or conceptual problem and a relatively painless solution to the problem.

In addition to the arguments given in the paper, there are two more general rules that should be applied to all political science data analyses.

First, we should concentrate on interpretable statistics. If the statistics are complicated, that is fine, as long as they can be translated into information that is meaningful to, and interpretable by, nonstatisticians.

Second, "getting a feel for data" is laudable, but presenting biased or incorrect results is not. Thus, we should try to use formal statistical models, about which much more is known. The problem with *ad hoc* solutions is that the same mistakes can occur in these as with formal statistics; however, we are much less likely to discover them. For example, political scientists are prone to doing a few cross-tabulations and arguing the point from there. Omitted variable bias, dichotomous dependent variables with linear models, and other specification issues are many times missed with this "method." What is often not realized is that these informal methods can usually be expressed in very simple formal statistical models. Their weaknesses then become immediately apparent.

If these two rules were followed—and adequate information were provided with which to assess the quantitative analyses— many future mistakes could be avoided.

Manuscript submitted 16 September 1985
Final manuscript received 18 November 1985

APPENDIX A
A PROOF THAT REGRESSION ON RESIDUAL (ROR) ESTIMATORS ARE BIASED

First, partition the coefficient vector as $b' = [b_1\ b_2]$ and the vector of independent variables as $X = [X_1\ X_2]$. Also let $Q = X'X$, $A = Q^{-1}X'$, and $e = My$ be the vector of residuals (where $M = I - XQX'$). Then b in the full regression, equation 2, is the least squares (LS) estimator, where $b = Ay$.

Now consider the regression on residual (ROR) estimator. First, let $Q_{ij} = X_i'X_j$ for $i = 1$, 2, $j = 1$, 2, $A_i = Q_{ii}^{-1}X_i'$ for $i = 1$, 2 and $M_1 = I - X_1Q_{11}^{-1}X_1'$. Then, calculate the coefficients and residuals with $\mathrm{b}_1^* = A_1y$ and $e_1 = M_1y$ from equation 3. b_1^* is the first ROR estimator. Then regress e_1 on the second set of explanatory variables X_2 and get $b_2^* = A_2e_1$ from equation 4, where, b_2^* is the second ROR estimator.

Now let $b^{*1} = [b_1^*\ b_2^*]$. I will first prove that $b \neq b^*$.

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \left[\begin{bmatrix} X_1' \\ X_2' \end{bmatrix} [X_1\ X_2] \right]^{-1} \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y$$

$$= \begin{bmatrix} Q_{11}^{-1}(I + Q_{12}(X_2'M_1X_2)^{-1}Q_{12}Q_{11}^{-1}) & -Q_{11}^{-1}Q_{12}(X_2'M_1X_2)^{-1} \\ -(X_2'M_1X_2)^{-1}Q_{21}Q_{11}^{-1} & (X_2'M_1X_2)^{-1} \end{bmatrix} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}$$

$$= \begin{bmatrix} b_1^* + Q_{11}^{-1}Q_{12}(X_2'M_1X_2)^{-1}X_2'M_1y \\ (x_2'M_1X_2)^{-1}X_2'M_1y \end{bmatrix}$$

$$= \begin{bmatrix} b_1^* + Q_{11}^{-1}Q_{12}(x_2'M_1X_2)^{-1}X_2'M_1(X_1b_1 + X_2b_2 + e) \\ (x_2'M_1X_2)^{-1}X_2M_1(X_2b_2^* + e_2) \end{bmatrix}$$

substituting from equations 2 and 4:

$$= \begin{bmatrix} b_1^* + A_1X_2b_2 \\ (X_2'M_1X_2)^{-1}(X_2'X_2)b_2^* \end{bmatrix} \neq \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix}.$$

Then, rearranging terms and taking expected values, we have:

$$E(b^*) = \begin{bmatrix} \beta_1 - A_1X_2\beta_2 \\ (X_2'X_2)^{-1}(X_2'M_1X_2)\beta_2 \end{bmatrix}$$

Thus, both b_1^* and b_2^* are biased. The former represents standard omitted variable bias.[19] There are also two special cases. If X_1 and X_2 are orthogonal (i.e., $X_1'X_2 = 0$), then $b^* = b$. Also,

[19] It is easy to see from this formulation that an omitted variable bias exists only when both (1) the sample coefficients (A_1X_2) resulting from the regression of the omitted variable on the included variables are nonzero and (2) the parameter on the omitted variable (β_2) is nonzero (i.e., has some influence on y). There is no bias if either one, or thier product, is zero.

when $b_2 = 0$, then $b_1 = b_1^*$. Finally, when $\beta_2 = 0$, $E(b_1^*) = \beta_1$. A similar proof can be found in Goldberger (1961). Furthermore, Goldberger and Jochems (1961) have shown for the bivariate case, and Achen (1978) for the multivariate case, that b_2^* is an underestimate of b_2.

APPENDIX B
THE RELATIONSHIP BETWEEN REGRESSION AND ANALYSIS OF VARIANCE

Note first that for the estimates of the model in equation 13, the following equalities hold:

$$\sum_{i=1}^{n} x = n_2 - n_1$$

$$\sum_{i=1}^{n} x^2 = n_2 + n_1$$

$$\sum_{i=1}^{n} y = n_2\bar{y}_2 + n_1\bar{y}_1$$

$$\sum_{i=1}^{n} xy = n_2\bar{y}_2 - n_1\bar{y}_1$$

Now, expressing b_0 and b_1 in terms of $\bar{y}_1$ and $\bar{y}_2$:

$$b_1 = \frac{n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2}$$

$$= \frac{(n_2 + n_1)(n_2\bar{y}_2 - n_1\bar{y}_1) - (n_2 - n_1)(n_2\bar{y}_2 + n_1\bar{y}_1)}{(n_2 + n_1)^2 - (n_2 - n_1)^2}$$

$$= \frac{\bar{y}_2 - \bar{y}_1}{2}$$

Also,

$$b_0 = \bar{y} - b_1\bar{x}$$

$$= \frac{n_2\bar{y}_2 + n_1\bar{y}_1}{(n_2 + n_1)} - \frac{(\bar{y}_2 - \bar{y}_1)(n_2 - n_1)}{2(n_2 + n_1)}$$

$$= \frac{\bar{y}_2 + \bar{y}_1}{2}$$

REFERENCES

Achen, Christopher H. 1977. Measuring representation: Perils of the correlation coefficient. *American Journal of Political Science,* 21 (November):805–15.

———. 1978. On the bias in stepwise least squares. Unpublished manuscript.

———. 1979. The bias in normal vote estimates. *Political Methodology,* 6(3): 343–56.

———. 1982. *Interpreting and using regression.* Beverly Hills: Sage.

Blalock, Hubert M. 1967a. Causal inferences, closed populations, and measures of association. *American Political Science Review,* 61 (March):130–36.

———. 1967b. Path coefficients *versus* regression coefficients. *American Journal of Sociology,* 72 (May):675–76.

Christoffersson, Anders. 1975. Factor analysis of dichotomized variables. *Psychometrika,* 40(1):5–32.

Friedman, Stanford B., and Sheridan Phillips. 1981. What's the difference? Pediatric residents and their inaccurate concepts regarding statistics. *Pediatrics,* 68(5):644–46.

Goldberger, Arthur S. 1961. Stepwise least squares: Residual analysis and specification error. *Journal of the American Statistical Association,* 56 (December):998–1000.

Goldberger, Arthur S., and D. B. Jochems. 1961. Note on stepwise least squares. *Journal of the American Statistical Association,* 56 (March):105–10.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica,* 37 (July):424–38.

Gurel, Lee. 1968. Statistical sense and nonsense. *International Journal of Psychiatry,* 6(2):127–31.

Hargens, Lowell L. 1976. A note on standardized coefficients. *Sociological Methods and Research,* 5 (November):247–56.

Hendry, David. 1980. Econometrics—alchemy or science? *Economica,* 47:387–406.

Huff, Darrell. 1954. *How to lie with statistics.* New York: Norton.

Kim, Jae-On, and G. Donald Ferree, Jr. 1981. Standardization in causal analysis. *Sociological Methods and Research,* 10 (November):187–210.

Kim, Jae-On, and Charles W. Mueller. 1976. Standardized and unstandardized coefficients in causal analysis. *Sociological Methods and Research,* 4 (May):423–38.

King, Gary. Forthcoming. 1986. Political parties and foreign policy: A structuralist approach. *Political Psychology.*

King, Gary, and Gerald Benjamin. 1985. The stability of party identification among U.S. senators and representatives. Paper presented at the annual meeting of the American Political Science Association, New Orleans.

Kritzer, Herbert M. 1978a. Analyzing contingency tables by weighted least squares: An alternative to the Goodman approach. *Political Methodology,* 5(4):277–326.

———. 1978b. The workshop: An introduction to multivariate contingency table analysis. *American Journal of Political Science,* 22 (February):187–226.

Leamer, Edward E. 1983a. Let's take the con out of econometrics. *American Economic Review,* 73 (March):31–44.

———. 1983b. Model choice and specification analysis. In Z. Griliches and M. D. Intriligator, eds., *Handbook of econometrics. Vol. I,* New York: North-Holland.

———. 1985. Sensitivity analyses would help. *American Economic Review,* 75 (June):308–13.

Lewis-Beck, Michael S. 1978. Stepwise regression: A caution. *Political Methodology,* 5(2): 213–40.

Sims, Christopher A. 1980. Macroeconomics and reality. *Econometrica,* 48 (January):1–48.

Smith, Kim. 1983. Tests of significance: Some frequent misunderstandings. *American Journal of Orthopsychiatry,* 53(2):315–21.

Tufte, Edward R. 1974. *Data analysis for politics and policy.* Englewood Cliffs: Prentice-Hall.