CONCEPTS OF RELATIVE IMPORTANCE

WILLIAM KRUSKAL
UNIVERSITY OF CHICAGO

How might one interpret the relative importance of independent variables, causes, or determiners when a dependent variable depends on those determiners together with chance? Such questions arise throughout science, technology, and national life. The paper attemps to clarify and critically describe a number of approaches to the problems of understanding relative importance.

Keywords: RELATIVE IMPORTANCE, REGRESSION, VARIANCE REDUCTION.

1. INTRODUCTION.

When two or more influences -call them variables, independent variates, causes, or whatever- affect a dependent variable, there is natural interest in understanding the relative importance or influence of the separate independent variables upon the dependent variable. A much-discussed example in the United States is the work of my colleague the sociologist James S. Coleman in his study of the acquisition of knowledge by school-children. His major conclusion -and I simplify for clarity of exposition- was that school facilities are of less importance than aspects of family background. See Mosteller, Frederick and Moynihan, Daniel P. /9/.

Another sociological colleague, W.J. Wilson /15/ studies the relative importance of race and of economic class in determining the well-being of Black Americans. Wilson concludes that race used to be the major determiner, but that economic class is now more important. He has been embroiled in considerable controversy around that theme, I think in part because of ambiguity in the meaning of relative importance.

The question of relative importance arises in public health settings. For example, a recent review in Science asks about the relative importance for cancer incidence of ocupation as against other factors (like parental health and smoking behavior). Another area of current interest is industrial productivity: how to compare the relative importance of education, work force morale, allocation of capital funds, etc. Indeed, considerations of relative importance arise so frequently, in so many contexts, and in such varied terminologies, that I am surprised by the relatively small attention paid to the topic in the statistical literature.

There are two general motivations for looking at relative importance; I might call them the technological and the scientific. The first, the technological, comes from the desire to change things effectively and economically; what should we attend to first in trying to reduce cancer deaths, improve education, maintain our systems of highways, increase productivity growth, etc.

The scientific approach is that of striving for basic understanding without special concern for immediate uses. Which variables should we examine in our next experiment or survey...since we never have the resources to examine all? how do the determiners of social class in a given country at a given era compare with each other in importance? What grammatical and semantic influences appear

- William Kruskal - University of Chicago - Dep. of Statistics - 5734 University Avenue
Chicago Illinois 60637 - U.S.A.

- Article rebut el novembre del 1983.
- Aquest treball es una ponencia invitada al "First Catalan International Symposium on Statistics" a Barcelona del 23 al 24 de setembre de 1983.

in the formation of pidgin or creole languages, and with what weights? What are the relative importance of sun, moon, and other planets in affecting the motion of the earth; and which may be reasonably disregarded at least for particular purposes?

Of course these two approaches overlap, but it is useful to bear both in mind. We want to avoid the kind of vagueness expressed in such traditional rhetorical questions as the relative importance of heredity and environment. That suggests a variant of a well-known Zen question: What are the relative importances of the right and the left had when clapping?

Ideas of relative importance enter in all scientific and non-scientific thought insofar as we simply cannot consider many variables at once. We are typically forced to reduce the variables under consideration to a small number, so that -usually tacitly- we behave as though most possible variables may be disregarded. As you know, there is a statistical literature that considers circumstances in which we start with an intermediate number of variables, for example, in a regresion study, and then reduce the number via a statistical algorithm.

The earliest statistical paper I know on relative importance is by R.H. Hooker and G.U. Yule /8/. They look at the production and exports of Indian wheat as determiners of the price of wheat in England. They arrive at the ratio of standardized regression coefficients, an approach suggested by many other later authors, for example by the widely used Snedecor and Cochran textbook /11/. In the Hooker-Yule paper, they find that the two independent variables (1) are highly correlated and (2) have nearly equal standardized regression coefficients. Both of these present recurrent problems, especially the first.

For if the independent variates were stochastically independent, or at least non-correlated,one might have a natural linear decomposition of the variance of the dependent variable. That independence among the independent variables is, however, rare except for special situations in which the scientist can provide it. (Be careful of confusion because of different senses of the term "independent".)

A paper by Frederick Williams and Frederick Mosteller /14/ is another early treatment of interest. They dealt with a sample of people cross classified into five economic strata and five educational strata. For each of the 25 cells thus formed, they look at the number of people in their sample who say "Yes" to a dichotomous opinion question. They then compute and compare two chi-square-like statistics that measure deviations from estimated expected counts under the hypotheses, respectively, of no economic stratum effect and no educational stratum effect. As they say, the statistics they use are arbitrary, and it is difficult to interpret them.

2. SIMPLEST CASES: ARE THEY PARADIGMATIC?

Perhaps the simplest case one can imagine is represented by elementary addition,

$$Y = X_1 + X_2,$$

where X_1, X_2, the independent variables, are the lengths of two adjacent bars in a mechanism, and Y, the dependent variable is the length of the assembly of the two bars. One might think of X_1 as representing the length of a randomly chosen bar from a bin of nominally identical bars of steel, and X_2 as the corresponding length of a brass bar from another bin. We might suppose the two choices stochastically independent. That fully specifies this tiny model, except for the distributions of X_1 and X_2; once we have those the model is fully given, and the next question is what are we trying to do or understand. The distribution of $Y = X_1 + X_2$ is in principle determinable at this point. We might want to know what the probability is that Y lies within prespecified limits, or we might be interested in the tails of the Y distribution, i.e., in the largest and smallest values. Is there any sense in which we can speak of the relative importance of X_1 and X_2? Several. Most simply we may note that

$$\text{Var } Y = \text{Var } X_1 + \text{Var } X_2$$

so that the ratio of variance, or of standard deviations might make sense as a measure of relative importance. Or one might be led to bring in the costs of tighter acceptance levels for X_1 as compared with X_2 if we are pointing towards lowering the variance of Y. On the other hand, variance per se may be irrelevant.

In some cases we might measure X_1 and then, by stratification of X_2, by machining the brass bar or otherwise, arrange matters so that $Y = X_1 + X_2$ is less variable. That is one way of introducing dependence between X_1 and X_2 to solve a problem...of course a price.

The simple additive model with which we began might also be used to represent a school test whose score is the sum of scores from two subtests. Here one would generally expect dependence, and full specification would require knowing the joint distribution of the X's. The comparative variances of the X's do not per se seem relevant. Possibly the correlations of X_i with Y are worth looking at, but I do not see any generally useful interpretation of them.

Note that in the above models, X_1 and X_2 are equally important, in the sense that a change in either of a given amount is equally reflected in Y. Thus there is no necessary connection between such a first moment-like approach to relative importance, and second-moment-like approaches via variances and correlation coefficients. Of course one may easily imagine variant models, e.g.,

$$Y = X_1 + 2X_2$$

where the effect of X_2 is doubled by mechanical linkage in the first example, or where the two parts of the test, in the second example, have different weights.

Notice that in these cases the model is wholly known, and there is no sampling or parameter estimation. Yet it is by no means clear even here how to regard relative importance. In my opinion, it is essential to treat the question first with a known model; time enough for the complications of sampling and estimation.

Some comentators think that the whole question of relative importance is itself unimportant, wholly ambiguous, and irrelevant. I do not agree because discussions of relative importance are so ubiquitous.

3. TABLE MODELS.

There is another very simple kind of model to discuss, for example

U.S. Death Rates per 100,000 from Accidents and Violence. 1979 (Rounded)

	Men	Women
Black	153	42
White	99	37

Source: U.S. Statistical Abstract 1982-3, p.79.

Is there any reasonable way of measuring the relative importance of race as against sex in this striking, poignant table, which shows higher rates for Blacks and for men, but by no means additively.

Similar tables arise frequently, often with such headings as

Region of Country

South North

or

Education of Parents

No parental college experience

At least one parent with some college

We might try some simpler set of hypothetical little tables

A

10	10
10	10

B

6	10
6	10

C

6	6
10	10

D

6	10
8	12

In A neither row nor column has an effect. In B, row has no effect but column does; vice versa in C. In D, a table with zero interaction, it seems clear that row has less effect than column: 2 units vs. 4. On the other hand maybe row and column are on totally disparate scales or represent dichotomies that could be sharpened. For example, Black vs. White is one kind of dichotomy; less than 21 years of age vs. 21 or older is quite different.In the latter case, the split could readily be put at other ages, and -more important- one might look at extremes, e.g., less than 15 vs. 30 or older, thus perhaps sharpening apparent effects.

Return to the accident death rate table and ask how we might approach it,

153	42	97.5
99	37	68.0
126.0	39.5	82.75

where the marginal numbers show averages. Here the difference between row averages is 97.5 - 68.0 = 29.5; between column averages 86.5. So from one viewpoint, column (sex) is considerably more important than row. Of course many other approaches could be taken, for example, looking at geometric averages.

4. BACK TO MOTIVATION. IMPORTANCE IN NATIONAL LIFE.

Issues related to relative importance are often central in major social and pclitical arguments. What are the primary causes of poverty? Of disease? Of war? What about crime? Some think that poverty is the major cause of crime. Others ascribe it to poor education. Still others speak of weakening of religious belief as a major cause of crime. One's point of view on such questions is inevitably intertwined with ideological and philosophical questions.

I do not have the courage or ability to attack the question of relative importance at that level of social interaction and public rhetoric, but I mention the question in passing as part of my motivation for interest in the area.

5. SLOPE-RELATED MEASURES.

In regression-like circumstances, where a dependent variable or its expected value,is expressed as a linear combination of independent variables

$$EY = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p,$$

possibly with an initial constant, β_0, we have a variety of models and measures of relative importance. Let me mention at first, and then discard to limbo, one approach to relative importance, that of significance tests. Some treatments test the separate null hypotheses that the β_i are zero and then assert that the greater the statistical significance (i.e., the smaller the observed tail probability, or P value), the greater the importance. That seems to me a serious mix-up of statistical and real significance and I mention the approach only to set it to one side. From now on we suppose that all parameters, and other aspects of distributions are known; there are enough problems then to keep us busy.

A linear model like the above is, of course, only a partial specification. It can be fleshed out to a full model in many ways. Of these, the two most common are what I call (1) the true multivariate model and (2) the fixed constants model. We shall deal only with these.

The true multivariate model, as I use the term, supposes that Y and the p x_i's have a joint distribution, possibly multivariate normal, in which the conditional expectation of Y given the x_i 's is linear. (Note that this model as flatly stated precludes polynomial or other forms in which the x_i 's are functionally connected.) We attend to the β_i 's, and we may at first simply accept as measures of importance the absolute values of the β_i 's (or perhaps their squares). This may make sense when the x_i 's are all variables of the same kind on a common scale, for example, temperatures in degree Celsius at 5 positions in a furnace, with Y a characteristic of the resulting product. If, however, some x_i 's are temperatures and others are amounts of catalyst or rates of liquid flow, then it hardly makes sense to look at the coefficients themselves. After all, each of those coefficients gives the rate of change per unit change in the independent variable, with everything else held fixed.
Indeed simply changing the scale from degrees Celsius to degrees Fahrenheit changes the coefficient correspondingly.

As a partial solution to this problem, some suggest standardizing by standard deviations, so that we look at $\beta_i\sqrt{\sigma_{ii}}$ or $\beta_i\sqrt{\sigma_{ii}}/\sqrt{\sigma_{yy}}$, where the σ 's are variances of x_i and Y. The latter are often called the standardized regression coefficients. These standardized forms are the regression coefficients if we scale the x_i 's (and Y) in terms of their standard deviations, i.e., look at the expectation of $Y/\sqrt{\sigma_{yy}}$ as a linear combination of $x_i/\sqrt{\sigma_{ii}}$. (In practice, we would also typically center about μ_i, the expectation of x_i, but that is not immediately relevant).

This device takes care of scale but in a possibly crude way. In particular, the dispersions of the x_i may be irrelevant to future applications. There is no general reason for supposing that marginal variances are invariant.

Another popular device for taking care of scale, especially when the x_i 's are inherently positive, as for some economic variables, is to say that we compare the effects on EY of (say) a one percent change in the independent variables. One percent of what? Presumably of expectation. So this means that we compare the quantities $\beta_i\mu_i$, or their absolute values. A difficulty with this is dependence upon origin: shifting from x_1 to $x_1 + 100$ would change relative importance without changing anything intrinsic.

Other difficulties with all the above are

No account is taken of trouble or cost in changing the x_i 's.

Does it make sense to talk of changing one x_i with the others held constant? Indeed it is precisely because the x_i come to us in linked or dependent form that we have a real problem.

Dispersion structure generally is not adequately examined (although, of course, the β_i in ordinary multivariate analysis are functions of the covariance matrix).

Attempts to apply these approaches to the fixed-constant model are bedevilled by the apparent meaninglessness there of talking about one variable at a time, unless the fixed constants are arranged to provide orthogonality.

6. VARIANCE REDUCTION MEASURES.

One naturally thinks early on (in the multivariate case) of using the squared correlations coefficients ρ^2_{yi} between Y and x_i to examine importance. A major reason is that ρ^2_{yi} measures the variance of x_i as a linear predictor of Y, and that $\sigma_{yy}(1-\rho^2_{yi})$ is the remaining variance of Y after removing linear prediction by x_i. Thus we might look at the separate marginal correlations as measures of relative importance.

The big problem with that approach is that it pays no attention to covariances among the x_i. An x_i may have small correlation with Y by itself, but may permit excellent prediction of Y with another x. Consider the clasic example

$$x_1 = Y + U$$

$$x_2 = U$$

where U, an error term, is uncorrelated with Y. Here, taking Var U = 1,

$$\rho_{y1} = \sqrt{\frac{\sigma_{yy}}{\sigma_{yy}+1}} \qquad \rho_{y2} = 0$$

If $\sigma_{yy} = .01$, $\rho_{y1} \cong .1$. Yet Y is perfectly predicted from x_1 and x_2; just subtract x_2 from x_1.

If there is a relevant ordering among the x 's, for example, an ordering of chronology or causation, then it may make sense to speak of the proportion of variance "explained by" the first x, the percent of the <u>remaining</u> variance "explained by" the second x, etc. (At least in principle, one might average over the various possible orderings.)

Two general concerns for any of these variance approaches are (1) does it always make sense to use variance as a measure of dispersion in this context?, and (2) terms like "explained" by may be misleading in suggesting non-existent causation.

7. COSTS.

I have mentioned briefly the possibility of bringing costs into the calculations, costs of making changes in an x_i or even of measuring an x_i. Costs of reducing variability may also be relevant. Brief discussions of costs in this context may be found in Williams/13/ and Carlborg /3/.

8. FINAL COMMENTS.

I have not returned to table models, the other general class, beyond my earlier remarks. They form an important class, but I have little to say about them now beyond standard discussions of main effects, interactions, and problems of interpretation when interactions are present.

My aim has been to draw attention to a kind of statistical problem that has received thus far inadequate attention. I hope that I have persuaded some of you to work on it.

A concluding cautionary note. In much writing about relative importance, the concept used is not explicit; with luck one can work back and see what the concept was, but in other cases it remains a mystery. I am concerned about robustness: in some instances conclusions about relative importance may stay much the same as one shifts among various concepts, but in other instances the conclusion may depend crucially on which concept is used. As theorists, practitioners, and critics of statistics, I hope that we become more sensitive to this problem.

9. BIBLIOGRAPHY.

/1/ ACHEN, CHRISTOPHER H.: "Interpreting and Using Regression". Beverly Hills: Sage. See esp. pp. 68-77. (1982).

/2/ BLALOCK, HUBERT M., Jr.: "Evaluating the relative importance of variables", American Sociological Review 26, 866-874. (1961).

/3/ CARLBORG, FRANK WILLIAM,: "A procedure for selecting independent variables in multiple regression", Ph.D. dissertation, University of Chicago. (1964).

/4/ FREEMAN, RICHARD B.: "Black economic progress after 1964: who has gained and why?" Pp. 247-294 in Sherwin Rosen (Ed.) Studies in Labor Markets. Chicago, Ill,: University of Chicago Press. National Bureau of Economic Research. (1981).

/5/ GREEN, PAUL E., CARROLL, J. DOUGLAS AND DeSARBO, WAYNE S.: "A new measure of predictor variable importance in multiple regression", Journal of Marketing Research, 15, 356-360, (1978).

/6/ HEDGES, LARRY V., AND OLKIN, INGRAM,:"The asymptotic distribution of communality components", Psychometrika 46, 31-336.(1981).

/7/ HOCKING, R.R.: "Developments in linear regression methodology", Technometrics 25 (1983) 219-230. Discussions 230-249.

/8/ HOOKER, R.H., AND YULE, G.U.: "Note on estimating the relative influence of two variables upon a third", Journal of the Royal Statistical Society 69, 197-200 (1906).

/9/ MOSTELLER, FREDERICK AND MOYNIHAN, DANIEL P.: "On Equality of Educational Opportunity", New York: Random House. (1972).

/10/ PEDHAZUR, ELAZAR J.: "Analytic methods in studies of educational effects", Review of Research in Education 3, 243-286.(1975).

/11/ SNEDECOR, GEORGE W., AND COCHRAN, WILLIAM G.: "Statistical Methods, seventh edition. Ames, Iowa: Iowa State University Press. See pp. 357-358, "Relative importance of different X-variables". (1980).

/12/ TUKEY, JOHN W.: "Causation, regresion, and path analysis", pp. 35-66 in Oscar Kempthorne et al (Eds.). Statistics and Mathematics in Biology. Ames, Iowa: Iowa State College Press. (1954).

/13/ WILLIAMS, EVAN J.: "Postcript to "Linear Hypotheses: Regression"", pp. 537-541 of William H. Kruskal and Judith M. Tanur (Eds.). International Encyclopedia of Statistics. New York: Free Press. (1978).

/14/ WILLIAMS, FREDERICK, AND MOSTELLER, FREDERICK,: "Education and economic status as determinants of opinion", Pp. 195-208 of Hadley Cantril (Ed.) Gauging-Public Opinion. Princeton, NJ: Princeton University Press. (1947).

/15/ WILSON, WILLIAM J.: "The Declining Significance of Race: Blacks and Changing American Institutions", Chicago: University of Chicago Press. Second edition. (1981).