Relative importance of explanatory variables: An annotated bibliography

David Firth
Nuffield College, Oxford OX1 1NF, UK


In the list below, an item with no annotation is one which I know contains some methodological discussion on relative importance, but which I have not yet seen. DF

References


‘Who are the most important people in the world? What is the most important part of an automobile? Questions like these are apt to be answered with another question: Important for what? Without a criterion for importance, the inquiries are meaningless.’ Goes on to distinguish three notions of importance, namely ‘theoretical importance’ (measured by $\beta_x$), ‘level importance’ (measured by $\beta_x\mu_x$) and ‘dispersion importance’ (measured by $\beta_x\sigma_x/\sigma_y$). Of the last, Achen says: ‘although almost no one is substantively interested in it, many social scientists use it as their sole importance measure.’ He suggests standardization by the (supposed fixed) range of a variable, rather than by its s.d., to achieve comparability across samples.


Suggests use of partial rather than marginal standard deviations when standardizing regression coefficients, an idea mentioned also in Healy (1990).


Points out erroneous reasoning in some silly published comparisons drawn from stepwise regression output.


Criticizes comparison of standardized regression coefficients for the familiar reason that they depend on the $x$ variances, not just on the relationship between the $x$’s and $y$. 
Statistics, Uppsala University.

Bring, J. (1996). A geometric approach to compare variables in a regression model. The

Shows how to interpret geometrically the standardized coefficients, studentized
coefficients, and decompositions of $R^2$ in multiple regression. Gives an example
in which the decomposition supported by Pratt (1987) seems unappealing.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative

Reviews various standard approaches. Then says ‘Most of the approaches re-
viewed...lead to conclusions...not invariant to subset selection.’ ... ‘Two re-
searchers who are interested in the relative importance of $x_1$ and $x_2$ may reach
different conclusions depending on the other predictors they include in their mod-
els. This situation reflects, in part, a true state of affairs induced by the nature
of the relationships between the different predictors involved. However, it would
be to everyone’s advantage to have a method of determining importance that is
invariant under all subset selections.’

Budescu’s method says that $x_1$ dominates $x_2$ if its partial correlation with $y$ is
greater regardless of the variables (from among those available — whatever that
means!) that are controlled for: the result is typically a partial ordering of the $x$
variables.

London.

Especially Section 4.5. Distinguish between ‘internal’ and ‘external’ standardiza-
tion, where internal means standardization determined mainly or entirely from
the data under analysis. Careful discussion of the usual kind of (internal) stan-
dardization and its limitations for interpretation.


Amongst other comments on this issue, Darlington is critical of the method
later supported strongly in Pratt (1987), calling it ‘meaningless’—though without
much real justification.


Statistician, 44:260.

A dismissive comment on Kruskal and Majors (1989). ‘I think...only unsophisti-
cated people try to make such assessments [of relative importance]’

Proposes a decomposition of $R^2$, different from that of Pratt (1987) (to which Genizi mildly objects because its components can be negative valued). Of the Pratt axioms, Genizi focuses on ‘orthogonal compatibility’, meaning that for a set of regressors partitioned into mutually uncorrelated subsets, the sum of components within any subset should be $R^2$ for that subset. Genizi’s proposed decomposition is based on a specially constructed orthonormal basis for the space of all regressors.


Criticizes Herrnstein and Murray’s assertions on relative importance of IQ and SES, which rest on comparison of standardized regression coefficients in logit models. ‘Standardization in the manner of HM—essentially using “beta weights”—has been a common practice in sociology, psychology and education. Yet it is very rarely encountered in economics ... is worse than useless ... yields misleading inferences.’ Similar criticism of measures based on variance explained. If relative importance is to have public policy meaning, the units of measurement must be money: ‘It would be meaningful to say that IQ is more important than SES if spending the [fixed] sum on IQ improvement rather than SES improvement were to yield a larger expected change in some outcome of interest.’


‘The usual methods of standardizing coefficients...distort the assessment of effects because they confound the effect of a risk factor with the standard deviations of the factor and the disease.’ Key point is lack of comparability between studies of the same effects but where the standard deviations are different. Critical remarks also about use of ‘variance explained’ in causal analysis as being irrelevant (e.g., to caseload, where the attributable fraction is more relevant). ‘In summary, standardized regression coefficients...have no meaningful biologic or public health interpretation as measures of effect.’


A very thoughtful essay. Says that ‘the sizes of effects are best measured experimentally, but their relative importance in practice must take the population frequencies into account.’ Defends standardization of coefficients thus: ‘Suppose we take two successive items from the population and get their $x$-values. The differences we observe will on average be proportional to the standard deviations, so that it makes sense to equate such differences when assessing their effects.’ To compare effects conditional on other variables in the model, Healy suggests standardization using conditional dispersion, as later elaborated by Bring (1994a). He also makes a comment about interaction, which is potentially important but largely ignored in this literature.
Similar objections to those of Goldberger and Manski on HM’s measurement of the relative importance of IQ and SES for outcomes such as poverty: ‘...the appropriate economic measure of importance of two variables is the relevant marginal cost for achieving a given change in the outcome.’ ... ‘Any [relative importance] measure that is divorced from cost considerations is hard to interpret.’


Criticizes standardization of coefficients. Makes the rather strong statement that ‘Standardization does not add information. If there was no basis for comparison prior to standardization, then there is no basis for comparison after standardization.’


Ridicules the criterion which was elegantly justified later by Pratt (1987).


Corrected 1987 vol 41, p341, to acknowledge priority of Lindeman et al., who make essentially the same suggestion. A recipe for removing dependence on the order of introduction of variables when measuring proportionate reduction in $R^2$: just average over all possible orderings. In essence, re-invents the same suggestion made in Lindeman et al. (1980).


Tries to unravel what Hooker and Yule meant to measure relative importance by (necessary since their account is confusing). Concludes that they used $\beta_x \mu_x$, the ‘level importance’ favoured by economists.


A rather diffuse survey of published scientific articles with the words ‘relative importance’ or similar in the title. Dismay that 20% of the papers they found used statistical significance to measure importance.

Kruskal here moderates his view of the Pratt (1987) approach, which had previously ‘seemed arbitrary’, but now, following Pratt’s axiomatic justification, rates as ‘a development that merits careful study’.

Argues for comparison of effect coefficients (direct plus indirect effects), rather than regression coefficients or correlations, to measure relative importance of variables in path models.


Page 120, with example on pages 125-127: essentially the same idea as Kruskal (1987), averaging reductions of variance over all orderings. Priority acknowledged by Kruskal in his 1987 correction.


‘[In multiple regression] it is natural and not unusual to try to measure relative importance without explicitly introducing further specifics. In this spirit, we shall put forth some properties or axioms that one would expect of any general procedure and show that they lead to a unique measure.’ The result is to partition $R^2$ into contributions \( \{\beta_x \sigma_x \rho_{xy}\} \), with $\rho_{xy}$ the ordinary correlation between $x$ and $y$. Pratt also shows that the competing suggestion of Kruskal (1987) violates two of his axioms, most crucially that the relative importance it attaches to $x_1$ depends on the coordinate system used for $x_2, \ldots, x_p$.


The last of these three papers by Schemper uses an analogue (developed in the earlier two), for survival analysis, of proportion of variance explained. Ranks variables by both their ‘marginal’ and ‘partial’ contributions. Describes a bootstrap method for inference on the resultant ranking(s).

With, say, death as outcome, consider a logit model

$$\log \left( \frac{p_i}{1 - p_i} \right) = \alpha + \pi_i + \phi_i$$

where $$\pi_i = \sum \beta_k x_{ki}$$ and $$\phi_i = \sum \gamma_m h_{mi}$$ are respectively the total effect of patient characteristics and of hospital characteristics. Silber et al. define

$$\omega = \frac{\text{variance of } \pi_i}{\text{variance of } \phi_i}$$

as measure of relative contribution to variation in the response. Obviously not restricted to logistic regression.


A clear, concise treatment (in Section 13.7) of the main methods, namely standardized coefficients and decomposition of $$R^2$$. On the dependence of contributions to $$R^2$$ on the order in which variables are introduced: ‘[In some applications] there may be a rational way of deciding the order in which the X’s should be brought into the regression, so that their contributions to $$\sum y^2$$ add up to the correct combined contribution. In his studies of the variation in the yields of wheat grown continuously on the same plots for many years at Rothamsted, Fisher [in *The Design of Experiments*] postulated the sources of variation in the following order: (1) A steady increase or decrease in level of yield, measured by a linear regression on time; (2) other slow changes in yields through time, represented by a polynomial in time with terms in $$T^2, T^3, T^4, T^5$$; (3) the effect of total annual rainfall on the deviations of yields from the temporal trend; (4) the effect of the distribution of rainfall throughout the growing season on the deviations from the preceding regression.’


I have not managed to find this yet. An online abstract by the same authors, for a conference paper titled ‘The Issue of Relative Importance of Variables in Statistical Decision Models’, reads as follows: ‘In a survey of scientific literature, Kruskal & Major (1989) have noted widespread misuse of significance testing for quantification of relative importance. We review some methods that are recently proposed for measuring relative importance of variables. We propose a Bayesian framework for dealing with the problem as a scientific basis for some recently suggested methods.’

*Simply a description of the multiplication of $\beta_x$ by $\sigma_x$ to achieve the same effect as standardization (to unit variance) of variable $x$. This is called ‘semistandardization’ because it does not eliminate the units of $y$.*


*These two papers build on Kruskal’s 1987 approach (averaging over all orders), but measuring each variable’s importance by the number of bits of information contributed by the variable to the whole. In essence, this is Kruskal’s method but with the averaging done on a logarithmic (base 2) scale.*


*Ridicules the same criterion so elegantly justified later by Pratt (1987).*


*Dismissive of variable importance assessments through partitioning of effects, except when variables are orthogonal. ‘In general the only realistic interpretation of a regression relation is that the dependent variable is subject to the combined effect of a number of variables. Any attempt to take the interpretation further [by partitioning] can lead only to misinterpretation and confusion.’*