MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 19, NUMBER 1, SUMMER 1992

A Case for Interpreting Regression Weights

Clifford E. Lunneborg
University of Washington (U.S.A.)
The Open University (U.K.)

Abstract

McNeil (1990) argues against interpreting estimated linear model parameters or weights, largely on the basis of the expected sample-to-sample variability in those estimates. In rebuttal it is noted that not to interpret model parameters is to ignore the strength of regression analysis. Appropriate regard may and should be given parameter uncertainty. But that is only part and parcel of parameter interpretation. Examples of linear model parameter interpretation are given.

Introduction

In a recent article in this journal McNeil (1990) writes, "Although most multiple regression texts argue against interpreting regression weights . . . some statistics text authors and researchers still want to place some sort of importance or meaning on the magnitude . . . of regression weights." Count me among them. Let me announce my loyalties even more strongly. I place not just "some sort of importance" on parameter interpretation; I regard interpretability as the central feature of a linear models approach to the analysis of both observational and experimental data. It is what holds us safe from the sterility of unrelieved null hypothesis testing. The case for interpretation I will base on a series of examples.

Examples of Linear Model Parameter Interpretation

Simple linear regression. Consider a simple (one explanatory variable) linear model. I'll assume the regression of Annual Income (in thousands of dollars) on Years of Education (in school years satisfactorily completed) is linear in some population of educated and employed individuals. So, we can write:

$$\text{E(Annual Income | Years of Education)} = \beta_0 + \beta_1 \text{ (Years of Education)} .$$

Our response variable (RV) and explanatory variable (EV) both possess metrics. So we have metrics for the regression slope and intercept as well:

β_0 : The expected Annual Income for an individual with 0 Years of Education. β_0 is some value in "thousands of dollars."

β_1 : The increase in expected Annual Income associated with an increase by one year in the number of Years of Education. β_1 is some value in "thousands of dollars per year."

Our slope or rate of change parameter has a simple and, I believe, very appealing interpretation. It tells us "how much" Education impacts Income. β_1 might be \$10/year or \$100/year or \$1,000/year or \$10,000/year.

Quite likely a Years of Education score of 0 is outside our range of interest; indeed, the distribution of Annual Income conditional on Years of Education being zero may be without any members. So, any interpretation of the intercept is uninteresting. We might anticipate this and choose to write our linear model in terms of a "Centered" Years of Education. In particular we might reduce Years of Education by a constant of 12 years giving

$$\text{E(Annual Income} \mid \text{Years of Education)} = \beta_0 + \beta_1 \text{(Years of Education} - 12) \,.$$

Our slope parameter has its same interpretation. β_0, though, is now the expected Annual Income (in thousands of dollars) for a high school graduate, a substantively more interesting quantity.

Given an appropriate sample from our population we can estimate these regression parameters. And, granted the satisfactoriness of our sampling assumptions, we can also know how much confidence to place in those estimates. It is my thesis that the point estimate of β_1 and its standard error are useful because we want to know how big the rate of change is, not because they allow us to "decide" between "rejecting" and "failing to reject" an hypothesis that β_1 is zero. My support for this borrows heavily from Tukey (1991); I but paraphrase.

Consider the following four possible confidence intervals for β_1, all, say 95% CIs:

Case A: [−\$10, \$8]

Case B: [−\$3,000, \$4,000]

Case C: [\$4,800, \$5,100]

Case D: [\$10, \$10,000]

Under Cases A or B we "fail to reject" the null hypothesis. But what a difference. Case A ought to tell us that the slope is flat; no question about it, expected Annual Income does not change with Years of Education. On the other hand, Case B ought to tell us that "we haven't the foggiest" whether Annual Income goes up, down or sideways! And, under either Case C or Case D we reject the null hypothesis. Yet Case D is rather like Case B in the lack of precision in our $\hat{\beta}_1$ while Case C allows us to say that an additional Year of Education increases the expected Annual Income by "almost exactly \$5,000." How we "decide the null hypothesis" is much less relevant than what we've learned about β_1.

McNeil (1990) inquires relative to the formula for the circumference of a circle,

$$\text{Circumference} = (\pi)\,(\text{Diameter})\,,$$

"... what does π mean? π is **simply** the weight which, when multiplied times the diameter, yields the circumference." I have added the emphasis. McNeil dismisses π too readily, as if all that were important about it is that it is some constant. But there is more to π. We think of it as a dimensionless number, but in the context of our Circumference equation it is a rate of change with a metric like in./in. or mm./mm. depending upon how we choose to measure Diameter. π is the amount by which the Circumference increases for a one unit increase in Diameter. Increase the Diameter of a circle by 1 inch and you increase its Circumference by (approximately) 3.14 inches. And the value of π has practical importance; it is a particular constant and it makes a day-to-day difference that it's value is what it is and not 5 nor 15 nor 1/5. Put another way, it is not sufficient to know that the Circumference of a circle is influenced by its Diameter or, equivalently, that π is greater than zero! As with mighty π, so too with our lowly βs.

Multiple Linear Regression. Now let's extend our Annual Income model by introducing a second EV, Parental Income (also measured in thousands of dollars per year). We write a model additive in the two EVs:

$$E(\text{Annual Income} \mid \text{Years of Education, Parental Income}) =$$
$$\beta_0 + \beta_1(\text{Years Education}) + \beta_2(\text{Parental Income})\,.$$

What interpretation do we give the β_1 of this model? It may be a little easier to see if we rewrite our linear model in the form

$$E(\text{Annual Income} \mid \text{Years of Education, Parental Income}) =$$
$$[\beta_0 + \beta_2(\text{Parental Income})] + \beta_1(\text{Years Education})\ .$$

The "slope" parameter, β_1, is still the expected increase in Annual Income for a one year increase in Years of Education (thousands of dollars per year of education). But, in this model the "intercept" takes different values depending upon Parental Income. So, our β_1 here has a **conditional** interpretation: The increase in expected Annual Income for a one year increase in Years of Education, for a **fixed** level of Parental Income.

Often an important question for modelled phenomena like this is whether the β_1 of our two EV model has essentially the same magnitude as the β_1 of our one EV model. Is the "influence" of Years of Education on Annual Income the same when we control for Parental Income (our conditional rate of change parameter) as when we ignore Parental Income? Note that the answer to this question has little to do with whether R^2 increases signficantly from the one to the two EV model. It has everything to do, of course, with the **substantive importance** of alternative values of β_1. As we have only estimates of the conditional and marginal rates of change we may seek refuge in the SEs. I emphasise, though, that the comparison is not a statistical but a substantive one.

Kleinbaum, Kupper & Muller (1988) discuss this comparison more fully, albeit under the somewhat pejorative title of "confounding." They take the position that where the two βs differ, we should prefer the conditional slope. That seems unwarranted. The two answer different questions. "What increase in Annual Income is expected for an additional Year of Education?" is one question. "What increase in Annual Income is expected for an additional Year of Education among those whose parents have identical annual incomes?" is a different question. We may be interested in whether the answers are the same or different, but to prefer one to the other is to pre-suppose the substantive question.

Moderated Regression Models. Our two EV model posits additive influences of Years of Education and Parental Income on our RV. More specifically, the "slope" parameter for Years of Education is assumed to be a constant, independent of Parental Income. We might have reason to doubt this assumption. It could be more realistic to assume that the Annual Income contribution of an extra year's education might itself be a function of Parental Income. In more conventional regression lingo we think that (a) Parental Income might **moderate** the influence of Years of Education or (b) we might need **separate slopes** for different Parental Income levels (as well as separate intercepts.)

The usual way of writing a moderated regression model is to allow the intercept and the regression slope of one EV (the moderated EV) each to be linear functions of a second EV (the moderator). In the present context we could write:

$$\begin{aligned}
&E(\text{Annual Income} \mid \text{Years of Education, Parental Income}) = \\
&[\beta_0 + \beta_2(\text{Parental Income})] + [\beta_1 + \beta_3(\text{Annual Income})](\text{Years Education}) = \\
&\beta_0 + \beta_1(\text{Years Education}) + \beta_2(\text{Parental Income}) + \beta_3(\text{YearsEducation} * \text{ParentalIncome}).
\end{aligned}$$

The bottom line above describes how we would "input" our regression model, introducing a product variable. It may be a good model to fit but it is quite unsuitable for interpretation. The slope parameter for the product variable, β_3, has (at least) two strikes against it: (1) Its metric is "thousands of dollars in Annual Income per unit of the product of Years Education and Parental Income." What a "unit" of the latter amounts to is not easy to grasp! (2) Even if we could come to terms with this complicated metric we are warned off interpreting β_3 because of its conditional nature. In effect, it assesses the contribution of the product variable when the other EVs in the model are held constant. But how can we think about a unit increase in the product of 2 EVs while each is held constant?

Fortunately, the intermediate expression above for our moderated regression model does invite interpretation. The regression slope for Years Education is given as:

$$[\beta_1 + \beta_3(\text{Parental Income})] .$$

This representation is faithful to the moderated regression assumption; the influence on Annual Income of an additional year of education varies with Parental Income. Given estimates of β_1 and β_3 the regression slope estimate is easily calculated for a selection of Parental Incomes of interest, say, \$20K, \$40K, \$80K, \$160K, etc. And, if our regression program provides (as it ought) the variance covariance matrix for the $\widehat{\beta}$s, it is also easy to calculate SEs for such linear forms of the $\widehat{\beta}$s as ($\widehat{\beta}_1$ + 20,000 $\widehat{\beta}_3$). Thus, CIs for the slope estimate at different values of the moderator can be provided.

Quadratic Regression. These ideas generalize to quadratic regression and, should the need ever arise to model RVs that double back on themselves in our design space, higher order polynomial regression. Say we thought Annual Income to be influenced quadratically by Years of Education as in this linear model:

$$E(\text{Annual Income} \mid \text{Years Education}) = \beta_0 + \beta_1(\text{Years Education}) + \beta_2(\text{Years Education})^2 .$$

(Years Education)^2 is not likely to vary independently of (Years Education) so there is little prospect of interpreting the two separate conditional slope estimates, β_1 and β_2.

However, if we rewrite the quadratic model as

$$E(\text{Annual Income} \mid \text{Years Education}) = \beta_0 + [\beta_1 + \beta_2(\text{Years Education})]\,(\text{Years Education})$$

there is a single slope to be estimated, but one which takes on different values depending upon where in the range of Years Education we want to estimate that slope. Quadratic regression is a special case of moderated regression; moderated and moderator variables are the same variable.

Interaction Models and Modular Models. One last example. We make Annual Income now a (probabilistic) function of two *categorical* EVs. We'll assume the population of interest to be college graduates and we are interested in modeling Annual Income (first year post-baccalaureate) as a function of Gender and Degree Major. For simplicity, the later takes only three "levels": Science, Social Science and Humanities. Allowing for the possibility of an interaction between Gender and Major we would likely begin modeling with a six parameter model. If our immediate goal were to test for a (significant) interaction this initial model might look like this:

$$E(\text{Annual Income} \mid \text{Gender, Major}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

where dummy variables have been employed as follows:

X_1: 0 for males, 1 for females.
X_2: 0 for Science or Humanities, 1 for Social Science
X_3: 0 for Science or Social Science, 1 for Humanities
X_4: the direct product, $X_1 * X_2$
X_5: the direct product, $X_1 * X_3$

The last two EVs can be thought of as "interaction variables" and the hypothesis of no interaction is tested by comparing the overall fits (R^2 or SS Residuals) of this model with one in which β_4 and β_5 are constrained to be zero (or, equivalently, X_4 and X_5 are "dropped" from the model.) If the difference in fits is non-significant we declare for the reduced, four parameter, additive model. We detected no interaction. Let's say, though, that the difference in fits was significant; either β_4 or β_5 or both are non-zero. Gender and Major do interact in influencing Annual Income. What do we do?

My belief is that we ought to do more than report that the interaction is significant or that the R^2 for the six parameter model is significantly larger than the R^2 for the four parameter model. We ought to "interpret" the interaction; how do Gender and Major interact? The βs for our two interaction variables, having as their metrics products of dummy variables, are not the best candidates for yielding up the desired interpretation. What works for me is to re-parameterize the interaction model into one with parameters that are themselves easily defined and give clear insight into the interaction.

First, what does the finding of an interaction mean, substantively? That the relative influence of the several Major levels on Annual Income is different for males than for females. Having learned this, it behooves us to model Major influence for males separately, somehow, from our modeling of Major influence for females. One way of looking at it is to say we want now to examine "simple" rather than "main" Major effects. That is facilitated by the re-parameterization to a **modular model**. The idea of the modular model is that it is equivalent to the interaction model (in numbers of parameters and fits) but consists of separate "modules" for each level of a categorical EV. (In the case of higher order interactions the modules may be for lower-order interaction "levels".) Modular models have been explicated primarily by writers on the use of weighted least squares in the analysis of categorical response data, e.g., Forthofer and Lehnen (1981) or Freeman (1987). However, they are equally useful in the linear modeling of a continuous RV. Here, we'd like separate modules for males and females.

Our modular model might look like this:

$$E(\text{Annual Income} \mid \text{Gender, Major}) = \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \beta_5 Z_5 + \beta_6 Z_6$$

where the linkage to our earlier dummy variables is as follows:

Z_1: X_1 (a 0/1 variable coding female)

Z_2: $1 - X_1$ (a 0/1 variable coding male)

Z_3: $Z_1 * X_2$ (a 0/1 variable coding female and social science)

Z_4: $Z_2 * X_2$ (a 0/1 variable coding male and social science)

Z_5: $Z_1 * X_3$ (a 0/1 variable coding female and humanities)

Z_6: $Z_2 * X_3$ (a 0/1 variable coding male and humanities)

Z_1, Z_3 and Z_5, together with their "weights", comprise the female module; β_1 is the "intercept" for the female module and β_3 and β_5 are the female module slopes for the dummy variables coding social science and humanities respectively. Correspondingly, the male module is based on Z_2, Z_4 and Z_6. Given our particular use of dummy variables, the intercepts evaluate to the expected Annual Incomes for (male and female) science graduates and the slopes to the differences between the expected Annual Incomes for either social science or humanities and those expected for science graduates (again, separately for males and females).

In fitting the modular model we obtain SEs for the six parameter estimates. While the presence of an interaction insures that we cannot have $\beta_3 = \beta_4$ and $\beta_5 = \beta_6$ simultaneously, we may be able to simplify the modular mode further, guided substantively by our re-defined parameters. The main point, though, is that the parameters of the modular model are directly interpretable and their estimates can be used to "explain" the interaction.

Discussion

I have tried, by example, to make the case for the directness and substantive importance of parameter interpretation in linear models. Why should it be controversial? I have not addressed that question but I think there are two issues involved. The first has to do with the stages of modeling, from model formulation through fitting and model comparison and on, perhaps, to model adoption. How we view a model and the relevance (or, indeed, acceptability) of parameter interpretation can depend upon the stage of modeling at which we are operating.

The second issue has to do with a contrast between phenomena that are thought to be wholly deterministic and those with an inescapable stochastic element. How we assess our success in modelling will depend on how much determinism we attribute to the phenomenon modeled.

Model Fit, Comparison and Interpretation. I ought make it clear that the model parameters whose interpretation concerns me are, for the most part, parameters in "accepted" or final models. I assume that we pursue our modeling with several alternative models in mind. These may all be pre-specified models, well rationalised in advance of any data collection or they may be models whose origin owes something to the "lay of the land" once we have it in sight. In either event, we are interested in identifying one or more of these alternatives as "better" than the others. Better, of course, must take into account the purpose for which we wish to find models.

Whatever our goal, however our alternative models are suggested to us, the path to an accepted model or models involves fitting several alternative models to our data and then comparing those fits. This fitting and comparing are done on statistical turf and parameter interpretation plays no role. Interpretation comes in after final, or, at least, promising, models have been identified. And it is necessary, in my view, if we are to do the best job of communicating our results. Interpretation, or the prospect of it, should also be kept in mind when we parameterise models. Every linear model permits of several alternative parameterizations, all providing the same R^2, the same fit to individual observations. We should choose one, our software willing, that will be natural to interpret later on. And, if our goals accommodate any degree of "model snooping", having parameters with simple interpretations makes it that much easier for us to see our way forward in model simplification or modification.

If we keep in mind where in the modeling process we are, we can make parameter interpretation work for us and not against us. I cannot believe that modeling progress is facilitated if the analyst is "blinded" as to the meaning of parameters **throughout the entire course** of modeling.

Deterministic and Stochastic Models. McNeil (1990) writes " . . . when one utilizes MLR one is taking the stance that behavior is complexly **determined** . . . The goal then is to account for the variation in the criterion by obtaining as high an R^2 as possible.." The emphasis is mine. In an appendix to the same article McNeil equates a "correct model" with one yielding an R^2 of 1.0. Both remarks suggest that he is modeling deterministic phenomena; given the right set of EVs, all of the response variability can be accounted for. Unquestionably, behavior **is complexly influenced** (if not wholly determined) and the search for a highest R^2 necessarily leads to models with very many EVs. And, indeed, in a model with 100, 200, perhaps more, intercorrelated EVs, parameter interpretation does become, at best, problematic.

Maximizing R^2 for deterministic responses is but one goal to be pursued with linear modeling. Let me suggest some alternatives.

(1) Not all behavioral, biological or social phenomena are deterministic. I mean that in two senses. First, there is the possibility of some inherent randomness; in principle we can never account for all of the variability in the free throw accuracy of NBA players. And, there are human limitations; in practice we shall never identify all of the EVs needed to account fully for the variability in the voting behavior of US state legislators. In either event, the "correct model" cannot extend beyond the EVs that are known to be relevant and will have an R^2 substantially less than 1.0.

(2) Even if we take the response to be deterministic, but complexly so, we often make scientific headway by considering, at one time, only a few of the many EVs which are known to be relevant. We seek to learn more about how some EV of interest influences a response. Several of my sketchy examples given earlier had a common theme; how does Years of Education influence Annual Income? Many, many factors other than Years of Education impact earnings. But, that's hardly the point if what I'm interested in is learning how Gender or Parental Income or College Major might moderate the influence of Years Education on Annual Income. If I sample randomly I need not worry overmuch about what else I might have put into my model.

(3) McNeil makes the very important point that the magnitude of an influence we detect for some EV in an observational study may be a poor guide to what happens when we attempt to manipulate that EV. That is a caveat to be heeded in the reporting of any observational study. Having said that, we can do worse in our search for potentially effective manipulations than to pay attention to the magnitudes of observational study influences. When I induce a student to remain in college another year I may not have increased her post-educational income by \$5,000 per year. Having noticed in an (hypothetical) observational study that, on average, each additional year of education was associated with that amount of additional income, however, suggests it is a manipulation worth trying, and evaluating.

I believe that a very great many, perhaps the substantial majority of, linear models in the biological and behavioral sciences are of these second and third kinds. They involve a limited number of EVs, often fewer than are known to be relevant to the RV. And, they address one or both of these questions: "How great is the influence, if any, of this EV?" and "How is the influence of my EV changed when I take these other things into account?" In neither case is the R^2 as relevant as the interpreted model parameters.

Parameter interpretation, far from being suspect, should be embraced by the multiple linear regression community. For appropriately parameterized models the parameters and their estimates provide natural measures of the magnitude of explanatory influences. Parameter interpretation is essential if we are to understand the meaningfulness (substantive significance) of an influence as well as its "presence" (statistical significance).

References

Forthofer, R. N. and Lehnen, R. G. (1981.) ***Public Program Analysis: A New Categorical Data Approach.*** Belmont, CA: Lifetime Learning Publications.

Freeman, D. H. (1987.) ***Applied Categorical Data Analysis.*** New York: Marcel Dekker.

Kleinbaum, D. G., Kupper, L. L. and Muller, K. E. (1988.) ***Applied Regression Analysis and Other Multivariable Methods.*** (2d. Ed.) Boston: PWS-Kent Publishing.

McNeil, K. (1990.) The case against interpreting regression weights. ***Multiple Linear Regression Viewpoints,*** 17 (2), 40-44.

Tukey, J. W. (1991.) The philosophy of multiple comparisons, ***Statistical Science,*** 6, 100-116.