Interpreting the substantive significance of multivariable regression coefficients

Jane E. Miller, Ph.D.[1]
[1]Research Professor, Institute for Health, Health Care Policy and Aging Research, Rutgers University, 30 College Avenue, New Brunswick NJ 08901, (732) 932-6730; fax (732) 932-6872, jmiller@ifh.rutgers.edu

Abstract: A critical objective for applications of multivariable regression analysis is evaluation of both substantive importance and statistical significance, yet many articles focus excessively on inferential statistical tests at the expense of substantive issues. I demonstrate approaches for writing clear sentences to interpret the real-world meaning of estimated coefficients from ordinary least squares regression, taking into account the type of independent variable and the distributions of the dependent and independent variables. After introducing "the Goldilocks principle" – that no one size contrast fits all variables – I use diverse examples to illustrate the importance of considering both the topic and data when evaluating substantive significance. Complementary use of prose, tables and charts to present both statistical and substantive significance are also covered.

Keywords: multivariable regression; statistical significance; substantive significance, writing

1. Introduction

Many papers that apply multivariable statistical methods to topics in the social sciences, health, or other fields are intended to shed light on a relationship among variables by testing a hypothesis derived from theory or previous empirical studies. For such papers, inferential statistics are a necessary tool for hypothesis testing, but another central consideration is the substantive significance of findings. A recent study by Ziliak and McCloskey (2004a) found that in applications of multivariable regression analysis in the economics literature, 80% of authors failed to distinguish between statistical significance and substantive importance. Their assertion sparked a debate about whether authors in fact conflate the two types of "significance," leading to a special issue of the *Journal of Socio-economics* devoted to the topic (e.g., Ziliak and McCloskey 2004b; Zellner 2004). That debate has identified the need for more guidelines on how best to present information on both statistical and substantive significance of regression results. This paper builds on work by Miller (2005) and Miller and Rodgers (2008) to provide and illustrate such guidelines.

1.1 What is substantive significance?

Substantive significance of an association between two variables is concerned with questions such as "how much does that association matter?" or "So what?" In various disciplines, substantive significance can be described as "clinically," or "economically," or "educationally" meaningful variation (Thompson 2004). In other words, substantive significance involves the real-world relevance of the statistical findings in the context of the specific topic under study. As an example, consider a recent study of how time spent playing video games relates to time spent on reading, homework, and other activities in a national sample of 1,400 U.S. adolescents (Cummings and Vandewater, 2007). In their model for boys, the estimated coefficient on gaming time is quite small (β = -0.04; s.e. = 0.01), which translates to a reduction of about two minutes in reading time for each hour spent video gaming. Although the estimated coefficient is *statistically* significant at the 0.01 level, it is so small that it isn't very *substantively* meaningful. In other words, that finding suggests that banning video games would not be a very effective way to increase reading time by a meaningful amount among adolescents, even if the association between gaming time and reading time is causal (see below).

An important part of writing a thorough description of multivariable regression results involves striking the right balance between presenting inferential statistical results and interpreting the substantive meaning of those results in the context of the particular research question. Both aspects of "significance" must be discussed because they address different analytic questions.

1.2 What questions does inferential statistics answer?

Statistical significance is an important aspect of an association between two variables. In multivariable regression, the statistical software calculates a test statistic (e.g., a *t*-statistic) based on the estimated coefficient (β) and its associated standard error, and then compares that test statistic against the critical value for the selected α-level (usually .05, corresponding to $p<.05$) and pertinent number of degrees of freedom. The *p*-value conveys the probability of falsely rejecting the null hypothesis (H_0: time spent gaming is not associated with time spent reading, or $\beta_{gaming} = 0$). Strictly speaking, $p<.05$ means that for a large sample such as that used in the gaming study, the estimated β_{gaming} is at least 1.96 times its standard error, that in the sample used to estimate the statistics, we have reached a conventionally accepted low level of probability of incorrectly rejecting no difference in the outcome of interest across the groups being compared.

Put differently, inferential statistics answers the question: "How likely would it be to obtain an estimated coefficient at least as large as that estimated for the sample, if in fact there is no effect of gaming on reading in the population from which the sample was drawn?" $p<.05$ corresponds to less than a 5% chance that there is no difference in time spent on homework between gamers and non-gamers given the estimated difference from the sample. In other words, if 100 samples were drawn from the U.S. population of adolescents, in only 5 of those samples would we observe a difference at least as large as that in the study sample if in that population, gamers and non-gamers spent equal amounts of time on homework (e.g., Agresti and Finlay, 1997).

1.3 What questions doesn't inferential statistics answer?

Although the above question is important, there are three other critical questions that inferential statistics does *not* answer but that are an important part of applications of multivariable analysis to real-world topics and data:

1.3.1 Causality

First, inferential statistics do not answer the question of whether the relationship is causal. Even the most basic statistics courses teach that "association does not imply causation," that two variables can be highly correlated without one causing the other. If two variables X and Y are associated because of their mutual association with another variable (Z), then the relationship between X and Y is confounded by Z. If Z completely explains the association between the other two variables, we say that the association between X and Y is spurious. Assessing causality is a key element of understanding a relationship between variables because non-causal associations should not be used to inform policy or program changes. If the independent variable X is not a cause of the dependent variable Y, then changing X will not result in a change in Y. Ruling out confounding or spurious associations is often a key reason why we estimate a multivariable model (Allison 1999; Miller 2005).

Certain types of study designs are better than others for assessing causality and determining direction of causation. Data from a randomized experiment in which the independent variable is manipulated and subsequent changes (or lack thereof) in the dependent variable are observed are considered the gold standard for assessing causality. Experiments and other "before and after" study designs provide more convincing evidence of a causal relationship than data where both the hypothesized cause and the hypothesized effect are measured at the same time (cross-sectional data). When random assignment isn't possible, "quasi-experimental" conditions can be simulated using multivariable regression (Allison 1999). See Morton, Hebel, and McCarter (2001), Schutt (2001), or other research methods textbook for an in-depth discussion of assessing causality.

1.3.2 Direction and magnitude

Second, inferential test results do not answer whether the effect has the expected sign. For example, the estimated coefficient of gaming time on homework time could be statistically significant but *positive*, in the opposite of the hypothesized direction. Third, inferential statistics do not convey whether the effect is big enough to matter in the real-world context. Recall the above example from Vandewater and Cummings (2007) that each hour spent gaming is associated with a 2-minute reduction in reading time. Is that enough to induce genuine concern on the part of parents or teachers?

This brief review of questions that are answered and unanswered by inferential statistics leads to the conclusion that one should not stop at $p<.05$ because it answers only part of what we want to know about the research question. In other words, inferential statistics are a necessary but not sufficient part of an application of statistical analysis to a real-world question. To understand the "importance" of an association between two variables, questions about substantive significance also need to be considered, including whether or not the association is causal, its direction (sign), and its magnitude.

2. Guidelines for substantive interpretation of regression coefficients

Although many statistics textbooks show how to assess and present statistical significance of regression results, few if any show how to interpret substantive significance, making it likely that researchers will overemphasize statistical significance at the expense of substantive interpretation. This paper shows how to achieve a balanced presentation of inferential statistics for formal hypothesis testing and interpretation of estimated regression coefficients in the context of the specific research question. The paper incorporates diverse examples in order to drive home the point that assessing substantive meaning of regression results requires an intimate acquaintance with both the topic and the data being used in the analysis.

2.1 Tools for presenting multivariable results

To present the information needed for both statistical and substantive interpretation of coefficients, use a combination of prose, tables, and charts in a paper about a multivariable analysis. These three tools have complementary strengths for presenting numbers: Tables and charts are good ways to systematically organize and present lots of detailed numeric information using a specific title, clear row and column labels, and footnotes. Tables are a place to put all the gory statistical information from multivariable models: Estimated coefficients, standard errors, test statistics, and p-values or symbols denoting statistical significance for each variable in the model, model goodness of fit statistics, sample size, and number of degrees of freedom. That information provides all the detail readers require to verify the assertions made in the text about the regression estimates.

See Miller (2005) or Morgan, Reichert and Harrison (2002) for guidelines on preparing well-organized, self-contained tables of statistical results.

Charts can be a useful tool for presenting the overall shape and size of a complex relationship among multiple variables. For example, an OLS model might be specified with a non-linear association between X_i and Y, such as with a quadratic or cubic function. Or, two independent variables X_1 and X_2 might interact in their association with Y (see below).

A few principles to keep in mind when writing a results section: Having reported the detailed numbers in tables, *interpret* them in the text, with an emphasis on the coefficients that relate directly to the main research question. An important way to convey the meaning of findings is to write in terms of the specific concepts involved in the specific research question rather than making generic references to "the dependent variable" or "the coefficient." A second way to improve clarity of presentation is to avoid using acronyms from your database – readers aren't going to be using your database, so they shouldn't have to flip back to the methods section in order to understand "alphabet soup" names in the tables, charts, or prose. Instead, use short phrases that refer to the concepts involved.

Third, use topic sentences to introduce the main questions to be addressed for each table or chart of results, and transition sentences to explain how one set of analyses relates to the next (Miller 2005; 2006). Fourth, incorporate units of measurement for each variable into the prose description as part of interpreting estimated coefficients. This will go a long way toward helping readers see the real-world implications of different effect sizes, and averting some pitfalls in interpreting coefficients. These points are illustrated with examples below.

2.1 Basics of interpreting coefficients

Here is a mantra to keep in mind when interpreting estimated coefficients: Report direction (AKA "sign"), magnitude, and statistical significance. Note that statistical significance is mentioned <u>*last*</u> in this mental checklist since many researchers focus on it as the main or even sole aspect of regression results, hence neglecting the substantive aspects of interpretation.

2.1.1 Direction

Writing about direction of association means conveying the sign of the relationship between the independent and dependent variables. Reporting direction is done differently for continuous than for categorical independent variables, because one can't discuss a "one-unit increase in gender," for example. For categorical variables, assess direction by determining which category has the higher value. As an example, if a model of income (in dollars) is specified with a dummy variable coded 1=male, and 0=female (the reference category), if $\beta>0$ males have higher income than females, whereas if $\beta<0$, the reverse is true. For continuous variables, β conveys whether the slope of the relationship between *X* and *Y* is rising, falling, or constant, such as whether the relationship between age (years) and income ($) is positive, negative, or level.

2.1.2 Magnitude

After conveying direction of association, the next step is to interpret the magnitude of association. Is the difference in earnings between males and females large or small? Is the slope of the association between age and income steep or shallow? The unstandardized coefficient from an OLS model measures the absolute difference in the dependent variable for a one-unit increase in the independent variable (*X*), where the effect size is in the original units of the dependent variable. For a continuous independent variable, β can be interpreted as the slope of the relationship between *X* and *Y*.

To see a concrete illustration of the abstract principle "specify direction and magnitude," examine the following series of "poor/better/best" versions of sentences. This approach, used throughout the paper, starts with samples of ineffective writing annotated to point out weaknesses, followed by concrete examples and explanations of improved presentation. Consider a model with mother's age as a predictor of birth weight, where birth weight is measured in grams and mother's age in years.

<u>Poor:</u> "Mother's age and child's birth weight are correlated (p<0.01)."

This sentence names the concepts (dependent and independent variables) involved and conveys statistical significance but omits direction or magnitude of the association.

<u>Better:</u> "As mother's age increases, her child's birth weight also increases (*p*<0.01)."

This version conveys concepts, direction and statistical significance but fails to report magnitude of the age/birth weight relationship.

<u>Best:</u> "For each additional year of mother's age at the time of her child's birth, the child's birth weight increases by 10.7 grams (*p*<0.01)."

Concepts, units, direction, magnitude, and statistical significance all in one simple straightforward sentence.

Categorical variables do not have units, so the estimated coefficient on a categorical predictor in an OLS model is interpreted as the absolute difference in the dependent variable for the category of interest compared to the reference category, holding constant the other independent variables in the model. Consider the example of gender as a predictor of birth weight.

<u>Poor:</u> "The β for 'BBBOY' is 116.1 with a s.e. of 12.3."

This sentence uses a cryptic acronym rather than naming the concept. It reports the same information as the table, but does not interpret it, leaving readers to calculate the test statistic & compare it against critical value. A results section should not be written like a statistics quiz, but rather should answer the underlying inferential statistical question

<u>Poor [version #2]</u>: "Boys weigh significantly more at birth than girls."

This version reports the concepts and direction of association but not the magnitude. Statistical significance is ambiguous: Is the term "significant" intended in the statistical sense or to describe a large difference?

Slightly better: "Gender is associated with a difference of 116.1 grams in birth weight (p<.01)."

This sentence conveys concepts, magnitude, and statistical significance but not direction: Was birth weight higher for boys or for girls?

Best: "At birth, boys weigh on average 116 grams more than girls (p<.01)."

Concepts, reference category, direction, magnitude, and statistical significance.

3. Charts to present complex patterns

A chart is often the best vehicle for portraying both direction and magnitude of an association because the relative heights of bars or slopes of curves are immediately apparent without readers having to do mental arithmetic from numbers in a table. These advantages make charts very effective tools for conveying the shape of a complex relationship, such as that involving more than one coefficient from a multivariable model. For such relationships, do the computations behind the scenes and then present the results of the calculations, generalizing the pattern rather than reporting all of the individual component coefficients or values. See Miller (2005) for more on how to generalize a pattern.

To illustrate, consider the set of estimated coefficients from a model of monthly earnings as a function of gender, marital status, shown in Table 1, with the other variables from the multivariable model listed in the footnote.

Table 1: Estimated coefficients for a model of monthly earnings (NT$) in Taiwan, 1992

Variable	*Coefficient*
Man	*3,205**
Married	*-1,595**
Interaction: Man and married	*4,771**

From Miller and Rodgers, 2008. * p<.05. Based on a multivariable model with controls for work experience, tenure, monthly hours, educational attainment, residence, and occupation characteristics.

The model includes an interaction between gender and marital status, so to see the overall pattern requires first adding together several coefficients to calculate how each gender/marital status combination relates to the reference category (unmarried women). For married men, the net effect involves both of the main effect terms and the interaction term: $\beta_{man} + \beta_{married} + \beta_{man\ and\ married}$, or 3,205 + (-1,595) + 4,771 = 6,381.

Poor: "The main effect of 'man' was 3,205 and the main effect of 'married' was -1,595, while the interaction term 'man and married' was 4,771 (all p<.05; Table 1)."

This description does not mention dependent variable (earnings), nor does it explain that the main effect and interaction terms must be considered together to calculate net effect of gender and marital status.

Poor [version #2]: "Gender and marital status interacted in their effects on earnings."

This version conveys that there is an interaction and names the concepts involved, but does not explain the direction or size of the relationship.

As a first improvement, Table 2 presents the net effects of the interaction, organized into one column for each gender and one row for each marital status, providing a neat grid with a cell for each possible combination of gender and marital status.

Table 2: Predicted difference in monthly earnings (NT$) by gender and marital status, Taiwan, 1992

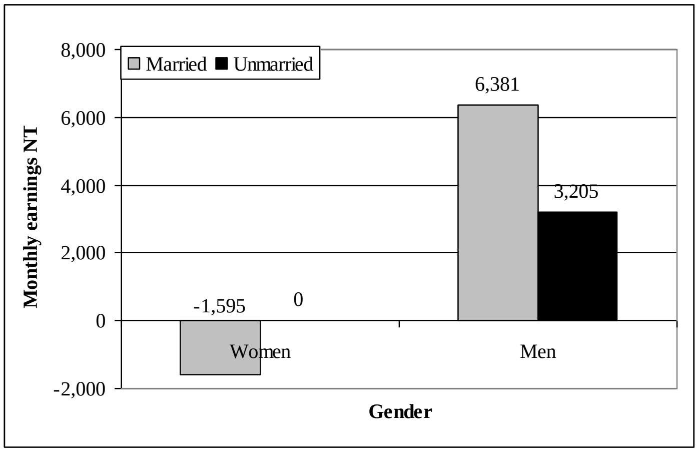
	Married	*Unmarried*
Women	*-1,595*	*0 (ref. category)*
Men	*6,381*	*3,205*

Slightly better: "The net effect of being a married man on earnings was NT$6,381 compared to unmarried women (p<.01; Table 2)."

This sentence reports the result of the calculation involving two main effects terms and the interaction term pertaining to married men, and correctly specifies the reference category. However, it does not interpret meaning of calculation or compare other gender and marital status combinations.

To improve the description further, replace Table 2 with Figure 1 to show the overall shape of the relationship among the four gender/marital status categories.

Figure 1: Predicted difference in monthly earnings (NT$) by gender and marital status, Taiwan, 1992



Source: Miller and Rodgers (1998). Based on a multivariable model with controls for work experience, tenure, monthly hours, educational attainment, residence, and occupation characteristics.

From this chart, it is easy to see both the direction and magnitude of the relationship between gender, marital status, and earnings. The most effective presentation would accompany Figure 1 with a text description such as:

> Best: "As shown in Figure 1, men earn more than women regardless of marital status, but the effect of marriage on earnings works in opposite directions for men than for women. Although marriage confers a substantial earnings advantage for men (NT$3,176 extra per month for married compared to unmarried men), it is associated with a sizeable deficit for women (NT$1,595 less per month for married compared to unmarried women)."
>
> *This description captures the overall shape of the earnings pattern among the four gender/marital status categories, including direction, size, and units.*

4. Pitfalls in interpreting coefficients

An important consideration when interpreting estimated coefficients from multivariable models is that not all coefficients are created equal. Simply scanning down a column of βs to find the largest numeric value that has asterisks next to it is *not* a good way to figure out which independent variable is the most important predictor in the model. Why? First, because variables come in different types (categorical and continuous), and second, because continuous variables can have wildly different levels, ranges, and distributions from one another. Recall that a β from an OLS regression estimates the change in the dependent variable for a one-unit increase in the independent variable, but for some variables a one-unit increase is not the most appropriate contrast.

4.1 Coefficients on categorical and continuous variables

One type of mistake in interpreting coefficients is to assess their relative "importance" by comparing them without regard to the kinds of variables involved. For instance, in the birth weight regression examined earlier, the estimated coefficient for "Mexican American" (-23.1) is bigger in absolute value than the estimated coefficient on "mother's age" (10.7). However, it does not make sense to conclude that ethnicity is a more important determinant of birth weight because one should not compare those coefficients directly. Mexican-American is a categorical variable, so its coefficient estimates the difference in mean birth weight for that ethnic group compared to non-Hispanic whites (the reference category). Mother's age, on the other hand, is a continuous variable so its coefficient estimates the difference in mean birth weight per one-year increase in her age, which can vary more than one unit (year) across cases (range in mother's age = 15 to 44 years). Even a two-year increase in mother's age is associated with birth weight difference nearly as large as the model estimated difference in birth weight between Mexican-American and non-Hispanic white infants.

4.2 The "Goldilocks problem"

No one size numeric contrast is universally suited to interpretation of all continuous independent variables. As Goldilocks discovered, some contrasts are too big, some contrasts are too small, and some contrasts are just right. Papa Bear, Mama Bear, and Baby Bear each had their own distinct requirements for the heights of their chairs and the dimensions of their beds. Identifying appropriate sized contrasts for interpreting estimated coefficients depends on the topics and variables involved in a particular research question. To figure out appropriate contrasts for a given set of variables, it is important to consider the level, range, and distribution of their values, and to evaluate them in the context of the subject and data at hand.

4.2.1 Too big

A common situation in which a one-unit contrast is too big involves an independent variable that measures a proportion. The valid range for a proportion is 0.0 to 1.0, so a 1.0 unit increase would constitute the entire theoretically possible range of such a variable, representing upper and lower bounds that rarely constitute a reasonably expectable change for most variables in the real world. Information about the distribution of the variable can help discern a more plausible contrast. Consider the example of monthly earnings in Taiwan as a function of the proportion of workers in an occupation that are women ("PWOW"). In the model estimated by Miller and Rodgers (2008), $\beta_{pwow} = -770.7$. Descriptive statistics for the sample used in the study show that the range of PWOW is from 0.01 to 0.95 with a standard deviation (σ) of 0.22. With those two pieces of information, one could calculate the effect of a one standard deviation increase in PWOW on earnings: $-770.7 * 0.22 = -170$, or a reduction of NT$170. This calculation helps convey expected change associated with a more plausibly sized contrast in PWOW in the context of the study sample.

In disciplines such as psychology, the problem of differing levels and distributions of variables is handled by using standardized coefficients, which estimate the effect of a one-standard-deviation increase in the independent variable on the dependent variable, where that effect is measured in standard deviation units of the dependent variable. Standardized coefficients provide a consistent metric in which to compare estimated coefficients on different variables, thus allowing assessment of the relative sizes of the associations of each independent variable with the dependent variable (Kachigan 1991). By moving out of the original units of measurement, however, they complicate intuitive interpretation of the substantive meaning, so it is often useful to explicate both the standardized and unstandardized coefficients to paint a complete picture of their real-world meaning.

4.2.2 Too small

As an example where a one-unit contrast is too small for the topic at hand, consider a study of biomedical risk factors for hospitalization based on multivariable hazards analysis of a large national survey of the U.S. (Miller et al 1998). In that study, the estimated relative risk among middle-aged men for systolic blood pressure was 1.004 ($p<.01$), meaning that each 1 mm Hg increase in systolic blood pressure was associated with a 0.4% increase in hospitalization risk. At first glance, this seems like a trivially small difference in hospitalization risk, but it is calculated for a very small increase in the independent variable.

In this instance, there are two issues contributing to the fact that a one-unit increase is too small for the topic. First is a measurement problem: It is very difficult to measure blood pressure within 1 mm Hg., so the precision of measurement implied by a one-unit increase is misleading. In other words, a wider increment is needed to suit the precision with which blood pressure can be measured in most clinical settings. Similar measurement or reporting issues occur with many other variables as well. How many Americans can recall their annual income from last year to the nearest dollar, for instance? The data were not collected at that level of detail, so it is not appropriate to report results to that level of precision.

The second issue is a specification problem: A one-unit increase in blood pressure is too small to be clinically meaningful. To find blood pressure contrasts that were large enough to be clinically relevant, the authors consulted with physicians and read the literature to identify well-established systolic blood pressure values for different levels of risk based on other outcomes like heart attacks & mortality. They then calculated relative risks of hospitalization for pairs of those values (shown in Table 3) to complement the detailed table of estimated relative risks per-unit increase in systolic blood pressure.

Table 3: Relative Risks of Hospitalization for Specified Values of Blood Pressure,
Men Aged 45-64 years, U.S. 1992 U.S. National Epidemiological Follow-Up Study

Variable	*Relative Risk*
Systolic blood pressure mm Hg	
140 versus 120	*1.09***
160 versus 120	*1.18***
180 versus 120	*1.28***

Source: Miller et al. 1998. ** p<.01
Based on a multivariable model with controls for age, race, exercise, diet, body mass index, laboratory tests, physical examination, and chronic health conditions.

4.2.3 Just right

To help identify a contrast that is the "right" size for variables in an analysis, it is essential to read the literature on the topic under study to find out what standards, cutoffs, or classification schemes are used. For instance, eligibility for the State Children's Health Insurance Program is based on multiples of the Federal Poverty Level (FPL). In New Jersey for instance, there are four SCHIP plan levels: Plan A (Medicaid) : <133% of FPL; Plan B: 133% to 150%; Plans C & D: 150% to 350% of FPL. If models of program take-up or retention are specified using those classifications of family income, it is easy to translate the results into policy or program recommendations. However, if the model is specified using purely empirical groupings such as quartiles or standard deviations of income, it is cumbersome if not impossible to show what the regression coefficients suggest about program outcomes as a function of income.

For topics that do not have established numeric standards, cutoffs, or classification schemes, consider different empirical criteria based on the distributions of the variables for identifying suitable numeric contrasts, such as standard deviations, interquartile range, or other plausible contrasts that fit the topic. When interpreting the results, explain the criteria used to select which size contrast to use for each independent variable, with reference to the literature, real-world issues such as program or policy design, and distributions of the variables in the data set.

4.3 Transformed variables

Another situation that complicates interpretation of estimated regression coefficients is when the model includes one or more transformed variables. Researchers transform their variables in a variety of ways such as taking logarithms or dividing by a constant (Gujarati 2002). For instance, in an OLS model of college grade point average (GPA) as a function of the student's own and roommate's SAT scores, Zimmerman (2003) divided SAT scores by 100 before entering those variables into the model. As a consequence, a one-unit increase in that transformed independent variable is equivalent to a 100-point increase in SAT. To put this question in substantive context: How likely is such an increase is in the real world? E.g., how many students are able to increase their math SAT score by 100 points between rounds? In Zimmerman's model, the β on own math SAT score was 0.092, so if an average student improved their math SAT by 20 points between their first & second attempts, that would correspond to a 0.018 point increase in GPA (=0.20 * 0.092). Calculations that bring the scale of transformed variables back to their original metrics are helpful for conveying the substantive importance of the coefficients involved. An exception is logarithmic transformations, which are often interpreted as elasticities by economists (Gujarati 2002).

4.4 Consider the range of the dependent variable

When interpreting the meaning of an estimated coefficient, it is also important to consider the range of the *dependent* variable. For instance, birth weight can range from roughly 500 grams to 5,000 grams for live-born infants. In contrast, grade point average (GPA) can only range from 0.0 (F) to 4.0 (A). As a consequence, a coefficient of given size will have very different substantive meaning for those two dependent variables. For birth weight, a β of 1.0 would be trivially small. For GPA, a β of 1.0 is one-fourth (25%) of the definitionally possible range.

5. Substantive and statistical significance in the discussion section

To round out a paper about an application of regression analysis to a topic in the social, behavioral, or other sciences, return to both substantive and statistical significance in the discussion section. Doing so brings readers back to the big picture of how the results of the multivariable model answer the original research question, summarizing causality, direction, and size as well as statistical significance of the main associations under study. These topics are treated differently in the discussion, with less numeric detail and statistical information than in the results section. In the discussion, describe statistical findings in words rather than numbers without repeating detailed standard errors, p-values, or test statistics that were reported in the results section.

5.1 Substantive significance in the discussion

Place findings back in the broader perspective of the original research question by addressing the following set of issues:

- Do the findings correspond to the hypothesis in terms of the direction and magnitude of the effect?
- Was the effect size attenuated when potential confounders or mediators were introduced into the model?
- What is the evidence for a causal relationship? If it isn't causal, what explains the association? If the association appears to be causal, what are the implications of the findings for policy, programs, etc.?
- How do the findings compare to those of other studies? Is there consensus or are there discrepancies across studies?

5.2 Statistical significance in the discussion

In the discussion, focus any mention of statistical test results on the *purpose* of those tests, addressing questions such as

- Was the estimated coefficient on the key independent variable robust to inclusion of other variables in the model, retaining its size and statistical significance? Or did its estimated coefficient become smaller and/or lose statistical significance when potential confounders or mediators were introduced into the multivariable model?
- Conversely, if a variable became statistically significant with inclusion of another variable or specification of an interaction effect (e.g., a suppressor effect), what does that change in statistical significance mean in terms of the underlying relationship among variables?

5.2.1 Lack of statistical significance

Although students often express great disappointment when some variable is not statistically significant in a multivariable model, a lack of statistical significance can be highly substantive significant. If theory or previous literature predicted a statistically significant association between the key independent variable and the dependent variable, a lack of statistical significance might be the most important finding of the analysis. In such cases, in the discussion section, the discussion section should include a systematic review of reasons for difference between the hypothesis and the statistical results.

There are several possible reasons why a coefficient might not be statistically significant. First, if the sample size is small, the coefficient might not reach conventional levels of statistical significance because the standard errors are too large. Other reasons for a lack of statistical significance relate to study design, measurement of variables, or model specification. These are key issues in an explication of the substantive significance of regression results. For instance, the analysis might involve an improved specification, such as one that includes a confounder or mediator that previous analyses didn't consider. A clear explanation of the theoretical roles of control variables and the logic behind estimating a series of nested models can help convey why inclusion of those variables contributes to the understanding of causal pathways linking the independent and dependent variables in the model.

Another reason for a lack of statistical significance when previous studies have found a variable to be statistically significant might be an improved measure of that variable. Or the analysis might have been based on data from a study design that is better suited for assessing causality than were previous studies. Finally, the analysis might involve a sample in which there is no association between the variables, even if a statistically significant association has been observed in a different time, place, or subgroup. Discuss these issues to help readers understand how to interpret the meaning of the findings about statistical as well as substantive significance.

6. Summary

This paper has presented a series of principles for achieving a balanced presentation of both statistical and substantive significance when writing about an application of multivariable regression to a real-world topic and data. Many researchers have been trained how to present statistical significance of estimated coefficients, so I have emphasized how to interpret the substantive meaning of those statistics. Some of these improvements can be made by using a judicious combination of prose, tables and charts to present regression findings. Use tables to report comprehensive, detailed inferential statistics, complemented with charts if needed to convey complex patterns. Use prose to ask and answer the research question, maintain a focus on the topic by writing about specific concepts and avoiding generic wording about "variables" and "coefficients."

To avoid common pitfalls in the interpretation of coefficients, get acquainted with the levels and distributions of the dependent and independent variables in the data used to estimate the model. Also read the literature on theory and empirical practice to identify whether there are numeric standard distributions or cutoffs for that topic and field, and to highlight the substantive issues behind the statistical analyses for that particular research question. Armed with that knowledge, researchers will be in a better position to design their model specifications to match the topic and data, and to select numeric contrast values that are both plausible and relevant for interpreting the real-world applications of the estimated regression results.

7. References:

Agresti, Alan, and Barbara Finlay. 1997. *Statistical Methods for the Social Sciences*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Allison, Paul D. 1999. *Multiple Regression: A Primer*. Thousand Oaks: Sage Publications.

Cummings, Hope M., and Elizabeth A. Vandewater, 2007. "Relation of Adolescent Video Game Play to Time Spent in Other Activities," *Archives of Pediatrics and Adolescent Medicine.* 161(7):684-689.

Gujarati, Damodar N. 2002. *Basic Econometrics*. 4th ed. New York: McGraw-Hill/Irwin.

Kachigan, Sam Kash. 1991. Multivariate Statistical Analysis: A Conceptual Introduction. 2nd ed. New York: Radius Press.

Miller, Jane E. 2005. *The Chicago Guide to Writing about Multivariate Analysis*. Chicago: University of Chicago Press.

Miller, Jane E., and Yana V. Rodgers, 2008. "Economic Importance and Statistical Significance: Guidelines for Communicating Empirical Research." *Feminist Economics*. 14(2):117-149.

Miller, Jane E., Louise B. Russell, Diane M. Davis, Edwin Milan, Jeffrey L. Carson, and William C. Taylor. 1998. "Biomedical Risk Factors for Hospitalization." *Medical Care* 36 (3): 411–21.

Morgan, Susan E., Tom Reichert, and Tyler R. Harrison. 2002. *From Numbers to Words: Reporting Statistical Results for the Social Sciences*. Boston: Allyn and Bacon.

Morton, Richard F., J. Richard Hebel, and Robert J. McCarter. 2001. *A Study Guide to Epidemiology and Biostatistics*. 5th ed. Gaithersburg, MD: Aspen Publishers.

Nicol, Adelheid A. M., and Penny M. Pexman. 1999. *Presenting Your Findings: A Practical Guide for Creating Tables*. Washington, DC: American Psychological Association.

Thompson, Bruce. 2004. "The 'Significance' Crisis in Psychology and Education." *Journal of Socio-Economics* 33(5): 607–13.

Zellner, Arnold. 2004. "To Test or Not to Test and If So, How? Comments on 'Size Matters.'" *Journal of Socio-Economics* 33(5): 581–86.

Ziliak, Stephen T. and Deirdre N. McCloskey. 2004a. "Size Matters: The Standard Error of Regressions in the American Economic Review." *Journal of Socio-Economics* 33(5): 527–46.

———. 2004b. "Significance Redux." *Journal of Socio-Economics* 33(5): 665–75.

Zimmerman, David A. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." *Review of Economics and Statistics*. 85(1):9-23.

Zveglich, Joseph E., Yana V. Rodgers, and William M. Rodgers. 1997. "The Persistence of Gender Earnings Inequality in Taiwan, 1978–1992." *Industrial and Labor Relations Review* 50(4): 594–609.