The Goldilocks Principle: Avoiding Pitfalls in Interpretation of Regression Coefficients

June 2008

Jane E. Miller, Ph.D.
Professor
Institute for Health, Health Care Policy and Aging Research, and
Edward J. Bloustein School of Planning and Public Policy

Mailing address
Institute for Health, Health Care Policy and Aging Research
Rutgers University
30 College Avenue
New Brunswick NJ 08901
Voicemail: (732) 932-6730; fax (732) 932-6872
Email: jmiller@ifh.rutgers.edu

Acknowledgements: I would like to thank Yana Rodgers and Joseph Zveglich for preparing the data used in this analysis, and Deborah Carr, Julie Phillips, and Louise Russell for comments.

Abstract

Regression coefficients estimate the effect of a one-unit increase in the independent variable X on the dependent variable Y. Although a one-unit increase is mathematically convenient, it is an arbitrary choice for interpreting the substantive significance of a coefficient. Depending on the type of variable and its level and distribution, a one-unit increase is too big for some independent variables, too small for others, and just right for yet others. This paper discusses theoretical and empirical criteria for choosing the right size contrast for each X_i to convey the size and shape of its relationship with Y. Transformation of variables and model specification are also considered. Example sentences are used to illustrate how to interpret coefficients in the context of the research question and data.

JEL Codes: A23; C10

New practitioners of regression techniques often commit one of two errors of interpretation of the estimated coefficients. One, they assess importance of different independent variables based solely on whether they are statistically significant. (How often have you seen students scan their printout or a published table of results for the asterisks often used to denote $p<.05$ to decide which ones to focus on?) Two, they assess which independent variables are the most important determinants of the dependent variable by comparing the sizes of their coefficients directly. Both approaches are problematic for interpreting the substantive meaning of coefficients.

A substantively significant association between two variables is defined as one that is economically, or educationally, or clinically meaningful for answering the particular research question (Miller 2005; Thompson 2004). In other words, it involves interpreting the relevance of the regression coefficient in the context of the topic under study. Although results of inferential statistical tests are an essential aspect of regression model results, *statistical* significance too often dominates the discussion of results, with little attention paid to *substantive* significance. The problem of conflating statistical and substantive significance has received much attention (e.g., Hoover and Siegler 2008; Ziliak and McCloskey 2008; Miller and Rodgers 2008), so it will not be addressed here.

Less widely discussed is the problem of assuming that regression coefficients can be compared directly to one another without consideration of the specific variables and metrics involved. Coefficients from ordinary least squares regression (OLS) regression estimate the effect of a one-unit increase in the independent variable (X_i) on the dependent variable (Y), where the change is measured in the units of the dependent variable (Gujarati 2002). Hence direct mathematical comparison of coefficients implicitly assumes that a one-unit increase in each independent variable is the pertinent contrast for that variable, whether income ($1), body weight (1 lb.), or work experience (1 year), for example.

But just as none of the chairs in The Three Bears' house fit all three of the bears comfortably, there is no one size of numeric contrast that will necessarily fit all variables in a regression model. A one-unit increase is a mathematically convenient but otherwise arbitrary choice for interpreting the substantive significance of a coefficient. Depending on the type of variable and its level and distribution, a one-unit increase will be too big for some independent variables, too small for other variables, and for yet others, just right.

Just as Goldilocks had to try out several chairs before arriving at the one that suited her, researchers should consider alternative contrasts before finding the one that suits each of their variables. And just as Goldilocks could have learned to anticipate how various chairs would fit without trying every one, researchers can learn key criteria for identifying appropriate contrasts for a given variable. An important task in teaching interpretation of regression coefficients is showing students how to choose the right size contrast for each independent variable to help convey the size and shape of its relationship with the dependent variable.

This paper illustrates these points with regression coefficients from an OLS model with several types of independent variables. Most examples are based on a regression model using data for female employees from Taiwan's 1992 Manpower Utilization Survey (N=7,944), a household survey that provides detailed information on individual workers' earnings, hours worked, educational attainment, tenure, job descriptors, and personal characteristics (Directorate-General of Budget, Accounting, and Statistics, Executive Yuan [DGBAS] 1992; Zveglich, Rodgers, and Rodgers 1997). Other economic topics and variables are included to demonstrate related points.

Fundamentals of writing about multivariate results

There are a few fundamental principles that can help improve communication of multivariate statistical findings. These principles are illustrated using a device called "poor/better/best" that starts with sentences that do not adhere to the principle, annotated to explain what is wrong with the sentence, followed by improved versions annotated to explain how they satisfy the principle. These examples show how to apply the abstract concepts about contrast choice and effective presentation of regression results to specific topics and coefficients.

To write clearly about multivariate coefficients, teach students to follow these simple principles, which are illustrated in the examples below:

- Report the detailed coefficients and standard errors for all variables in a table along with model goodness of fit statistics. This table provides readers with the full set of statistical information they need about model estimates without cluttering up the prose with numeric details.
- Write in terms of concepts, not variables, avoiding phrases that refer to "the coefficient" or "the dependent variable" without also identifying it.
- In prose, tables, and charts, substitute meaningful phrases for the acronyms used to abbreviate variables within the data set used for the analysis. Readers aren't likely to be using the same database, so why make them have to guess or look up abbreviations in the methods section?
- Embed the units of measurement for each continuous variable in the prose description of coefficients as well as in all tables and charts; for categorical variables, identify the reference category and name the categories being compared.
- Describe the direction (sign), magnitude, and statistical significance of the association between the independent and dependent variables rather than simply reporting the coefficient, standard error, and *p*-value, which readers can find in the table of detailed findings. See below for more on explaining magnitude of association, taking into account the type and scale of each variable.
- In the interest of conserving space use a phrase such as "*ceteris paribus*," "holding all else constant," or "controlling for other variables in the model" when interpreting the first coefficient from a multivariate model, and then do not repeat it for other coefficients from that model. A subheading such as "*Multivariate results*" also can be used to differentiate a section of bivariate or three-way findings from a section of multivariate results, averting the need to state that information for every coefficient.

Consider types of variables

The first consideration in assessing the suitability of a one-unit increase is the type of independent variable. Variables can be classified into one of two broad types: continuous and categorical. Continuous variables are measured in units such as years (e.g., age or date) or currency (e.g., earnings or price), including those that assume only integer values as well as those with decimal values. Categorical variables come in two types: Ordinal ("ordered") variables have categories that can be ranked according to the values of those categories, such as primary, secondary, and higher education. Nominal ("named") variables such as gender or marital status are classified into categories with no inherent order.

With continuous variables, the concept of a one-unit increase is consistent with how those variables are measured. For a continuous independent variable the unstandardized coefficient from an OLS regression is an estimate of the slope of the relationship between X_i and Y. For instance, the coefficient on number of children under age 15 estimates the marginal effect of a one-unit increase in X_i (an additional child) on Y (monthly earnings), holding constant all other variables in the model. In the earnings model shown in Table 1, $\beta_{nkids<15} = -486.95$.

Poor: "The coefficient on NKIDSLT15 was -487 (s.e. = 97.16)."

This sentence simply repeats information from the associated table without interpreting it. It also uses a confusing acronym for the variable name and fails to mention the dependent variable or its units.

Better: "In a model of earnings, β on number of children was -487 (p<.01)."

This version names the concepts measured by both the independent and dependent variables, but does not interpret the coefficient or specify the units of the dependent variable.

Best: "For each additional child under age 15 years, a woman's monthly earnings decreased by NT$487 (p<.01)."

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

For categorical variables, evaluating a one-unit increase is not meaningful (Chambliss and Schutt 2003). The numeric distance between categories of an ordinal variable such as education level cannot be assumed to be constant, so a "one-unit" increase is misleading. The values of nominal variables such as place of residence (urban versus rural) have no natural numeric order, so it makes no sense to say "as place of residence increases…" Consequently, the coefficient on a dummy or binary independent variable such as "urban" compares values of Y for the category of interest (urban, coded =1) to the reference category (rural, coded = 0). In Table 1, the coefficient on "urban" is interpreted as: "Women in urban areas earned on average NT$973 more per month than their rural counterparts."

Given these different interpretations, if a model specification includes both continuous and categorical independent variables, it does not make sense to compare those coefficients directly. For urban/rural (a categorical variable), the contrast is one category versus the other. For number of children (a continuous variable with a range of zero to six children in the Taiwan sample; Table 2), the contrast can vary more than one unit (child) across cases. As a consequence, although the coefficient on "urban" (β_{urban} = 973; Table 1) is larger than the coefficient on number of children ($\beta_{nkids<15}$ = -487), one cannot conclude that place of residence is a more important determinant of earnings than is number of children.

Examine range of values

As shown in the preceding example, the range of values assumed by a continuous independent variable provides important information for determining the size of a plausible contrast to use when assessing the meaning of its associated OLS coefficient. Topic alone isn't sufficient to determine plausibility of a numeric contrasts because range and scale of a given variable often differ substantially by context. For example, annual income would be far lower for data collected in the early 20th century than in the early 21st century; for a sample of teenaged, part-time workers than for a sample of their older, full-time counterparts; or for less developed than for more economically developed countries.

<u>When a one-unit change is too big</u>

If the entire theoretically possible range for an independent variable is 1.0 unit (or less), clearly using the estimated coefficient with its implied one-unit increase to assess the effect of a change in that variable is implausible. One example of a variable for which a one-unit increase is too large is the Gini coefficient, which measures inequality of distribution of income or wealth, with a value of 0.0 representing perfect equality and 1.0 complete inequality (Sen 1973). For such variables, a one-unit increase would encompass the full range from the lower to the upper bound of the variable rather than a more credible change in observed values, which might be only, say, 0.10.

Likewise, variables measured as proportions can range only from 0.0 to 1.0, as with the variable "proportion of workers in an occupation that are women" from the regression of women's monthly earnings (Table 1). Interpreting the coefficient on such a variable directly without regard to its observed range in the data will overstate the magnitude of the association with the dependent variable in most applied situations because a one-unit increase would constitute the entire theoretically possible range of such a variable.

A related caveat: Researchers are often sloppy in their labeling of variables measured in proportions, instead describing them as percentages. The percentage equivalent of a proportion is by definition 100 times as large, so readers need to know the actual scale of the variable used in the model in order to be able to interpret its coefficient correctly. Although for a variable measured as a proportion a one-unit increase is too large, for a variable measured as a percentage a one-unit increase may well be too small. Make sure the labeling of units is correct and consistent across data set, tables, charts, and prose.

When a one-unit change is too small

There are many independent variables for which a one-unit contrast is too small to matter given the range and scale of possible values. For example, Table 2 shows that in the Taiwan sample the average female employee worked just over 200 hours per month, so a one-hour increase in monthly hours worked is a trivial change. A more suitable change might be the difference between part-time and full-time work. Depending on the distribution of values, variables measuring percentage of a whole (e.g., percentage of the population that is poor) could suffer from this problem.

Scale changes and decimal system biases

For some research questions, a simple change of scale can help make a one-unit contrast in the independent variable align better with the research question, or shift the scale of the estimated coefficients to be more consistent with those of other variables in the model. Typically, those scale changes involve dividing or multiplying the original variable by a multiple of ten– a reflection of a type of bias that pervades cultures oriented to a decimal system. Working with multiples of ten facilitates the calculation of the new variable but a one-unit increase in the new variable might not correspond any better to the real-world context of that variable than did a one-unit increase in the original units. Put differently, although changing the scale of a variable by an order of magnitude or two is a mathematically convenient transformation, it is also arbitrary and in many cases unrelated to the topic or data under study.

Resolving the Goldilocks problem

For variables for which a one-unit contrast is either too large or too small to be realistic given the range and scale of values, there are two main criteria for choosing a pertinent contrast: *Theoretical* considerations for the concepts at hand, and *empirical* information for the analytic sample.

<u>Theoretical criteria for choosing a numeric contrast</u>

In many social science fields, there are thresholds defined by social policies (e.g., the income threshold used to define poverty in a given place and time) that facilitate a choice of a relevant and realistic contrast to apply to the associated coefficient. For instance, eligibility for the State Children's Health Insurance Program (SCHIP) is defined based on percentages of the federal poverty level (FPL; U.S. Department of Health and Human Services 2008). A sensible choice for illustrating the effects of differences in family income in a model of children's health insurance coverage would be to specify income as a percentage of the FPL and then compare outcomes for families with income at the poverty threshold (income = 100% of FPL) against families with income at 200% or 250% of the FPL, rather than using family income in dollars.

Alternatively, thresholds or classifications for a particular concept can be used to create categorical versions of the variable to be included in the model specification. For instance, in the analysis of women's earnings in Taiwan, highest education level has been classified into primary school or less; middle school; high school; and so forth as shown in Table 1. To determine the appropriate classifications for a given data set, it is important to examine the distribution of the original variable to make sure the categories make sense for that time, place, and group. For instance, in countries where few people attain a secondary education, categories should differentiate among lower levels of education.

Likely real-world changes in the value of a variable can also be used to identify useful contrasts. For example, legislation calls for the U.S. minimum wage to increase from $6.55 in 2008 to $7.25 per hour in 2009 (U.S. Department of Labor 2008), so it would make sense to apply a $0.70 increase to the coefficient on hourly wage in a model predicting poverty. Reading the academic and program or policy literature about the topic under study is a critical step for identifying pertinent-sized contrasts.

<u>Empirical criteria for choosing a numeric contrast</u>

Lacking theoretical guidelines, information about the distribution of the variable can help identify a useful contrast. For instance, a one-standard deviation increase in X can be used to assess a realistic change. In the 1992 Taiwan survey, one standard deviation in the proportion of workers in the occupation that are women (PWOW for short; σ_{PWOW}) = 0.22. Using that value, the earnings penalty from working in an occupation that has a plausibly higher value for the PWOW can be calculated by multiplying the contrast of 0.22 by -814.1 (β_{PWOW} from Model I, Table 1): -814.1 * 0.22 = -179.

Poor: “The effect of proportion of workers in an occupation who are women on earnings is NT$ -814.1.”

Comment: The distribution of values of PWOW in the Taiwan data suggests that a contrast of 1.0 unit is far too large for the empirical context. Evaluating the coefficient without consideration of its range substantially overstates its effect.

Better: “A one standard deviation increase in the proportion of workers in an occupation that are women decreases average monthly earnings by NT$ 179.”

Comment: By applying a size of contrast that fits the data used to estimate the regression model, this example provides a more realistic appraisal of the effect of PWOW on earnings.

In fields such as psychology and sociology, standardized coefficients are used to 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 (also known as the "beta coefficient" or "beta weight"; Allison 1999) adjust for the fact that some variables have a much larger standard deviation than others; hence a one-unit absolute increase means different things for different variables. They thus allow assessment of the relative sizes of the associations of each independent variable with the dependent variable. In order to preserve the intuitive understanding of the coefficient, complement reporting of standardized coefficients with an explanation of how the findings translate back to the original units of each variable; see Miller (2005) for examples.

Another empirical approach is to apply the interquartile range for X as a contrast. For instance, the 25th and the 75th percentiles for monthly hours worked are 191 hours and 208 hours, respectively, resulting in an interquartile range of 17. Multiplying that value by $\beta_{monthly}$ hours yields 17 * 36.47 = 619.99. When writing up the results, conduct such computations behind the scenes (with perhaps one example in the text multiplying the contrast by the coefficient to illustrate), and then report and interpret the result of that calculation in the text. "Women at the 75th percentile of monthly hours worked earned on average NT$ 620 more per month than those at the 25th percentile of hours worked."

<u>Consider the dependent variable, too</u>

So far, this paper has focused on the range and distribution of independent variables in regression models, but those same criteria should also be used to evaluate the dependent variable, Y. For instance, although an estimated coefficient of 1.0 is a trivially small effect in a model predicting monthly earnings (which ranges from NT$ 500 to NT$ 132,000 in the sample used here), the same coefficient would be quite substantial in a model predicting grade point average on the usual 4-point scale, where a 1-point increase is one-quarter of the possible range. As a consequence, it is important to consider the range, scale, and distribution of both the independent and dependent variables when interpreting the substantive meaning of regression coefficients.

Issues in variable creation and model specification

Many of these Goldilocks issues can be addressed by considering the scale of variables before specifying the model and transforming variables or altering the model specification at that point.

Transforming variables

These approaches include using model specifications in which (1) the values of the variable is divided by some scalar, e.g., income in $1,000s, or population in millions; (2) a smaller level of aggregation (weekly income instead of annual income, or population for smaller geographic regions) is used; or (3) the independent and/or dependent variables are logged. If variables are transformed in any of these ways, it is critical to mention the calculation and specify the units so that coefficients can be interpreted correctly. In tables of descriptive statistics, also report information for the untransformed version, such as income in dollars as well as ln (income).

Mention the system of measurement, units, and level of aggregation for all variables in the methods section, and then label them accordingly in all tables, charts, and prose. For instance, if a model is specified with annual income in $1,000s, report those units in the table of coefficients so that readers don't assume the coefficient is per $1 increase in monthly income.

Non-linear specifications

Up to this point, this paper has addressed only interpretation of coefficients for independent variables that have been specified as having a linear association with the dependent variable. In a linear relationship the effect on Y of a one-unit increase in X is constant across the range of X values. Interpretation of coefficients that involve non-linear specifications such as polynomial functions or logarithmic transformations of X is more complicated because the effect of a one-unit increase in X varies across its range of values. For instance, in the model shown in Table 1, the positive coefficient on the linear term ($\beta_{experience}$ = 417.9) coupled with the negative coefficient on the quadratic term ($\beta_{experience-squared/100}$ = -765.0) yields an earnings function that increases at a decreasing rate as work experience increases, holding other variables in the model constant. For example, moving from 1 to 2 years of work experience is estimated to increase monthly earnings by NT$ 395, whereas moving from 20 to 21 years of work experience is associated with an increase of only NT$ 104. For non-linear specifications, then, it is important to describe the overall shape of the relationship, including a couple of numeric contrasts that illustrate that shape. Graphs of such patterns can be invaluable for conveying complex patterns.

Specifications in which the independent or dependent variable, or both, are logged (not shown) yield coefficients that are interpreted in percentage rather than absolute terms. See Gujarati (2002) for additional explanation, Miller and Rodgers (2008) for example specifications and sentences to interpret these coefficients.

Summary and conclusion

This paper has sought to demonstrate that a "one-size-fits-all" approach to interpreting the substantive meaning of regression coefficients is misleading because variables come in different types, with different units of measurement and varying distributions. OLS regression coefficients provide estimates of the effect of a one-unit increase in an independent variable on the dependent variable, but for some independent variables a one-unit increase is too large for its range and scale, for some it is too small, and for still others, it is just right.

This Goldilocks problem can be minimized by careful attention to the units and distributions of variables during data preparation, model specification, and presentation of results. An essential first step is to become familiar with the variables under study by examining their distributions in the data set and reading the literature on the research question. That information should then be used to inform sensible decisions about how to define variables and specify the model, in some cases changing the scale of variables, creating a categorical version of a continuous variable, or using a non-linear specification for some variables.

The next step is to write a clear prose interpretation of the substantive meaning of the regression coefficients, using the principles outlined in the "Fundamentals" section. Simply reporting each coefficient increases the chances of making Goldilocks errors by encouraging readers to compare the coefficients to one another without consideration of variable type, range, or scale. To avert that problem, explain the criteria used to arrive at a numeric contrast that fits both the topic and the data. Then interpret the direction, magnitude, and statistical significance of a "right sized" contrast for each of the key variables in the model. By complementing a table of detailed model estimates with a prose description of what those findings mean, researchers can more effectively convey how the results of their multivariate analysis apply to the real world question they are analyzing.

References cited

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

Chambliss, D. F. and R. K. Schutt. 2003. M*aking Sense of the Social World: Methods of Investigation*. Thousand Oaks, CA: Sage Publications.

Directorate-General of Budget, Accounting, and Statistics, Executive Yuan (DGBAS). 1992. *Manpower Utilization Survey Database*. Taipei: Directorate-General of Budget, Accounting, and Statistics.

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

Hoover, K. and M. Siegler. 2008. "Sound and Fury: McCloskey and Significance Testing in Economics." *Journal of Economic Methodology*. 15(1):1-37.

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

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

Sen, A. 1973. *On Economic Inequality*. Oxford, England: Clarendon Press.

U.S. Department of Health and Human Services, 2008. "HHS Releases New Analysis of SCHIP Eligibility." Available online at http://www.hhs.gov/news/facts/schip.html. Accessed June 2008.

U.S. Department of Labor, 2008. "History of Federal Minimum Wage Rates under the Fair Labor Standards Act, 1938 – 2007." Available online at http://www.dol.gov/esa/minwage/chart.htm ; accessed April 2008.

Ziliak, S. T. and D. N. McCloskey. 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives.* Ann Arbor: University of Michigan Press.

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

Table 1 Ordinary least squares regression of monthly earnings (NT$), women aged 15–65 in Taiwan, 1992

	Unstdized. coeff.	Std. error
Intercept	628.52	621.08
Job characteristics		
Manager or supervisor	4,320.37**	345.50
Proportion women in occupation	-814.10*	359.16
Personal characteristics		
Live in urban area (rural)	973.20**	148.86
Married (unmarried)	2.76	221.74
Number of children <15 years	-486.95**	97.16
Productivity characteristics		
Highest education level attended		
(Primary school or less)		
Middle school	2,406.11**	278.78
High school	5,381.89**	333.89
Vocational school	5,715.84**	300.12
Junior college	11,178.87**	324.34
College or higher	17,302.42**	370.63
Potential years post school experience		
Experience	417.88**	23.89
$\text{Experience}^2/100$	-764.98**	48.35
Enterprise-specific tenure (years)		
Tenure	527.99**	36.88
$\text{Tenure}^2/100$	-419.44*	147.81
Monthly hours worked	36.47**	2.35
Number of observations (N)	7,944	
F statistic (df)	467.06 (15)**	
Adjusted R^2	0.47	

Notes: The data are from Taiwan's 1992 Manpower Utilization Survey, and the sample is restricted to all civilian women of working age who are non-farm, paid employees.

The variable "Proportion women in occupation" is the proportion of workers in an occupation who are women.

Reference category in parenthesis. * $p<0.05$; ** $p<0.01$

Source: DGBAS 1992

Table 2 Descriptive statistics, earnings sample, women aged 15–65 in Taiwan, 1992 (N=7,944)

	Minimum	Maximum	Mean	Median	Std. dev.
Monthly earnings at primary occupation (New Taiwan $ = NT$)	500	132,000	18,837	17,000	8,727
Monthly hours worked	13	537	202	208	31
Potential post school experience, years	0	59	14.8	12	12.4
Enterprise-specific tenure, years	0.1	42.0	4.5	2.8	5.1
Proportion women in occupation	0.01	0.95	0.58	0.56	0.22
Number of children < 15 years	0	6	0.64	0	1.03

Notes: The data are from Taiwan's 1992 Manpower Utilization Survey, and we restrict the sample to all civilian women of working age who are non-farm, paid employees.

The variable "Proportion women in occupation" is the proportion of workers in an occupation who are women.

Source: DGBAS 1992