

10. *The “Goldilocks Problem” in Multivariate Regression*

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. In a journal article in your field, circle all numeric contrasts.
 - a. Indicate whether in each instance the author provides enough information for you to assess whether it is a realistic difference or change for the research question and context.
 - b. Review the authors’ description and interpretation of coefficients on each continuous variable in their multivariate model, considering whether they
 - i. reported associated units for the dependent and independent variables
 - ii. stated the size of the contrast used to interpret the size of the coefficient
 - c. Evaluate whether different or additional size contrasts would be useful for the intended audience, considering
 - i. plausibility;
 - ii. real-world application;
 - iii. measurement issues.
 - d. Identify an audience that would be interested in different applications than the audience for whom the article is currently written, e.g., an applied rather than academic audience. Describe how you would select numeric contrasts to meet their interests.
2. Find a journal article that estimates an OLS model with some continuous and some categorical independent variables. Evaluate whether the authors explicate the coefficients in ways that differentiate those types of variables and their associated scales, using the criteria in chapter 10 of *Writing about Multivariate Analysis, 2nd Edition*. If not, rewrite the description to rectify those errors.
3. Find a journal article in your field about an application of an OLS model with standardized coefficients for at least two continuous independent variables.
 - a. Evaluate whether they discuss why estimate that specification, with reference to the distributions of the variables, the shape of their relationship, or theoretical reasons for their topic.

- b. Evaluate whether they interpret the coefficients in ways that clearly convey the scale and substantive importance of the respective variables in the model.
 - c. Evaluate whether the units of the statistical test information are consistent with the units of the standardized coefficients. If not, suggest a correct alternative for presenting statistical test results, using the guidelines in chapter 11.
4. Find a journal article in which the authors estimate models with one or more logarithmic specifications (log-lin, lin-log, or log-log). Review whether the authors
- a. Discuss why they estimate that type of specification, with reference to the distributions of the variables, the shape of their relationship, or theoretical reasons for their topic.
 - b. Interpret the coefficients in ways that explicate their units and the shape of the association with the dependent variable.
 - c. Rewrite their description of results to address any shortcomings you identified in parts a and b, using the guidelines in chapter 10 and the associated online materials.
5. Find a journal article that presents results of an OLS model involving a quadratic specification for an independent variable.
- a. Critique the description of the coefficient for that variable, using the criteria in chapter 10.
 - b. Rewrite the description to correct any shortcomings you identified in part a.

B. Applying Statistics and Writing

1. Calculate and graph the frequency distribution of a continuous independent variable using the highest possible level of detail (e.g., the smallest units for that variable available in your data).
- a. Name the shape of the distribution (e.g., normal, uniform, skewed).
 - b. Mark the cutpoints for the quartiles of that variable on the chart.
 - c. Mark ± 1 standard deviation (SD) and ± 2 SD on the chart.
 - d. Assess the appropriate scale of numeric contrasts for that variable given the precision with which it was collected.
 - e. Evaluate whether there is appreciable heaping in the reported values of that variable.
 - f. Referring to your answers to parts a through e, explain the criteria you will use to select appropriate values to contrast within your data as you illustrate model findings in your results section.

2. Answer the following questions using the graph you created in the preceding question.
 - a. If you wanted to use a categorical version of that independent variable in your model, what does the graph suggest might be empirically appropriate cutpoints between categories? Why?
 - b. Read the literature on the relationship between that independent variable and your dependent variable. Are there standard ways to classify the independent variable?
 - c. Are there policy-, program-, or other “practical” criteria related to your research question that suggest ways you might classify that variable?
 - d. Do the empirical cutoffs you identified in part a match the cutoffs you found for parts b and c? If not, explain which of these criteria you will use to classify your data and why they suit your intended audience.
 - e. Design a table or chart to contrast results obtained using the approaches to classifying your independent variable in parts a through c.

3. Complete the “Getting to Know Your Variables” assignment, available in the supplemental online materials.

4. Assess the appropriate scale of numeric contrasts for each of the variables in your analysis given the precision with which those data were collected.

5. Identify several continuous independent variables in your data that have different scales and distributions (e.g., one that is measured in proportions, others that range up to values in the thousands or higher). Investigate the various “Goldilocks solutions” described in chapter 10 of *Writing about Multivariate Analysis, 2nd Edition* for an OLS model involving those variables.
 - a. Transforming one or more variables.
 - b. Specifying a model with standardized coefficients.
 - c. Specifying one of the logarithmic specifications.
 - d. For the methods section of your paper, describe how you arrived at your preferred solution and how it affects the definitions of variables or model specifications.

6. Using data on the same variables as in question B.1 of the suggested course extension for chapter 9 (a continuous dependent variable, denoted Y_1 in the equations below, and a continuous independent variable denoted X_1), estimate the following variants of an OLS model. For each, write a sentence interpreting the value of β_1 , referring to the variables you have used and specifying the units using

the guidelines on pp. 221–23 of *Writing about Multivariate Analysis, 2nd Edition*.

- a. $Y_1 = \beta_0 + \beta_1 X_1$ (in the original, untransformed units of both the dependent and independent variables, but specifying standardized coefficients)
 - b. $\ln Y_1 = \beta_0 + \beta_1 X_1$ (a log-lin model)
 - c. $Y_1 = \beta_0 + \beta_1 \ln X_1$ (a lin-log model)
 - d. $\ln Y_1 = \beta_0 + \beta_1 \ln X_1$ (a double-log model)
7. Using the same variables as in the preceding question estimate an OLS model with a quadratic specification of X_1 : $Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$.
- a. Calculate the predicted value of Y_1 for selected values of X_1 that span its observed range in your data.
 - b. Consider whether increments other than a one-unit increase in X_1 are better suited to your research question and data, following the guidelines in chapter 10.
 - c. Create a chart to show the shape of the estimated relationship between Y_1 and X_1 , using the results from part a and following the guidelines in chapter 6.
 - d. Write a sentence to describe the relationship between Y_1 and X_1 across the observed range of X_1 in your data, using the calculations from parts a or b.
 - e. Optional: Use a spreadsheet to perform parts a and c, using the quadratic spreadsheet template available online or following the instructions in appendix D.

C. Revising

1. Critique a data and methods section you have previously written, considering each of the following and using the guidelines in chapter 10 of *Writing about Multivariate Analysis, 2nd Edition*:
 - a. Reporting of the units of measurement for all variables in your analysis.
 - b. Descriptions of the distributions of all continuous variables, and how those distributions affected the ways in which you specified those variables in your statistical models;
 - c. Description of the precision of measurement of your variables and the implications for how you analyzed those variables;
 - d. Explanation of the calculations and reasons for transformations you made to any of the variables, including references to standard transformations or classifications used in your field;
 - e. Description of your model specification and how it was affected by Goldilocks issues, including references to standard practice in your field;

- f. Revise the data and methods section to rectify any shortcomings you identified in parts a through e.
2. Critique a table of descriptive statistics you previously created, using the criteria in chapter 10 to evaluate the following elements:
 - a. Labeling of units (system of measurement, units, and scale) and categories for all variables, following the guidelines in chapter 4;
 - b. Labeling of all variables measured as proportions, percentages, or rates that correctly conveys their units and scale as used in the multivariate model specification;
 - c. Labeling of all transformed variables that correctly conveys their units and scale or categories as used in the multivariate model specification;
 - d. Pertinent measures of central tendency and distribution for each variable, given its level of measurement;
 - e. Revise the table to rectify any shortcomings you identified in parts a through d.
3. Critique a table of multivariate regression results you previously created, considering each of the following as explained in chapter 10:
 - a. Labeling of units and categories for all variables;
 - b. Labeling of all variables measured as proportions, percentages, or rates that correctly conveys their units and scale as used in the multivariate model specification;
 - c. Labeling of all transformed variables that correctly conveys their units and scale or categories as used in the multivariate model specification;
 - d. Title, row or column headings, or footnotes to convey the model specification (e.g., standardized or unstandardized coefficients, logarithmic specification);
 - e. Revise the table to rectify any shortcomings you identified in parts a through d.
4. Critique a description you have previously written about results of an OLS model with several continuous independent variables with different ranges and scales of values.
 - a. Evaluate whether you specified the size of the contrast used to interpret the coefficients for each continuous variable.
 - b. Consider whether a one-unit contrast was suited to the interpretation of the coefficients for each of those variables, based on the criteria in chapter 10.
 - c. Revise the description to rectify any shortcomings you identified in parts a and b.

5. Critique a description you have previously written about results of an OLS model with at least one continuous independent variable and one categorical dependent variable, considering whether you clearly conveyed
 - a. the nature of the contrast that suited each type of independent variable;
 - b. for ordinal variables, the substantive meaning of a one-unit increase (from one category to the next);
 - c. whether you directly compared the size of coefficients on categorical and continuous variables;
 - d. Revise the description to rectify any shortcomings you identified in parts a through c.

6. Critique a description you have previously written about a quadratic association between one of your independent variables and your dependent variable, using the guidelines in chapter 10.
 - a. Assess whether a chart would complement the narrative description. If so, create one, using the guidelines in chapter 6 and the spreadsheet template available online or the instructions in appendix D.
 - b. Revise the description to improve the shortcomings you found.

7. Critique and rewrite a description you have previously written about estimated coefficients from one or more types of logarithmic specifications, using the guidelines in chapter 10.

8. Critique and rewrite a description you have previously written about estimated coefficients from a model with standardized coefficients, using the guidelines in chapter 10.

9. Exchange revised drafts of the materials in questions C.1 through C.8 with someone writing about a different topic or data set. Peer-edit each other's work and revise according to the feedback you receive.