# 9

# Quantitative Comparisons for Multivariate Models

## SUGGESTED COURSE EXTENSIONS

### A. REVIEWING

1. Find a journal article in your field that presents results of an OLS model with at least one categorical independent variable and at least one continuous independent variable. Use the results to answer the following questions.
   a. Critique the description of the coefficient on a continuous independent variable in terms of direction, magnitude, statistical significance, and units, using the criteria in chapter 9 of *Writing about Multivariate Analysis*.
   b. Critique the description of the coefficient on a categorical independent variable using the criteria in chapter 9.
   c. Rewrite the descriptions of both coefficients to correct any problems you identified in parts a and b of this question.

2. Find a journal article that presents results of an OLS model with an interaction between a categorical independent variable and a continuous independent variable.
   a. Do the authors calculate and present the net effect of the interaction? If so, how? If not, suggest an appropriate tool for presenting the results for this research question and audience.
   b. Critique the authors' written description of the interaction using the criteria in chapters 9 and 13 of *Writing about Multivariate Analysis*.
   c. Rewrite the description of the interaction to correct any shortcomings you identified in parts a and b of this question.

3. Find a journal article that presents results of an OLS model with an interaction between two categorical independent variables.
   a. Do the authors calculate and present the net effect of the interaction? If so, how? If not, suggest an appropriate tool for presenting those results for this research question and audience.
   b. Critique the description of that interaction using the criteria in chapters 9 and 13.
   c. Rewrite the description of the interaction to correct any shortcomings you identified in parts a and b.

4. Find a journal article that presents results of an OLS model with standardized coefficients.
   a. Critique the description of the coefficient for one variable, using the criteria in chapter 9.

b. Rewrite the description to correct any shortcomings you identified in part a.

5. Find a journal article that presents results of an OLS model with a lin-log or log-lin specification.
   a. Critique the description of results, including correct interpretation and units of effect size, using the criteria in chapter 9.
   b. Rewrite the description to correct any shortcomings you identified in part a.

6. Find a journal article that presents results of a logistic regression of a binary dependent variable, with at least one categorical independent variable and at least one continuous independent variable. Use the results to answer the following questions.
   a. Do they report log-odds or odds ratios? If odds ratios, do they interpret them in terms of multiples of odds or multiples of risk?
   b. Critique the description of the effect size for a continuous independent variable in terms of direction, magnitude, statistical significance, and units, using the criteria in chapter 9.
   c. Critique the description of the effect size for a categorical independent variable.
   d. Rewrite the descriptions to correct any shortcomings you identified in parts b and c.

## B. APPLYING STATISTICS AND WRITING

Notes: For the "applying statistics" questions, use variables from your own data to substitute for $Y_1$, $Y_2$, $X_1$, *DUMMY,* and *CATEGVAR* in the models described below. For example, suppose you want to examine factors that predict income. You might use income in dollars as a continuous dependent variable ($Y_1$), educational attainment in years as a continuous independent variable ($X_1$), gender as a binary independent variable (*DUMMY*), and residence (urban/suburban/rural) as a multicategory independent variable (*CATEGVAR).* If you wanted to study factors that predict poverty, you might use poverty status (poor/nonpoor) as a categorical dependent variable ($Y_2$), with the same set of independent variables.

If possible, choose variables that are part of an ongoing research project. Save the computer output from the models you estimate in questions B.1 through B.6 for use in the exercises for chapter 10.

1. Using data on a continuous dependent variable (denoted $Y_1$ in the equations below) and a continuous independent variable (denoted $X_1$ in the equations below), estimate the following variants of an OLS model. For each, write a sentence interpreting the value of $\beta_1$, referring to the variables you have used and specifying the units using the guidelines in chapter 9 of *Writing about Multivariate Analysis*.
   a. $Y_1 = \beta_0 + \beta_1 X_1$ (in the original, untransformed units of both the dependent and independent variables, with unstandardized coefficients)

b. $Y_1 = \beta_0 + \beta_1 X_1$ (in the original, untransformed units of both the dependent and independent variables, but specifying standardized coefficients)
c. $\ln Y_1 = \beta_0 + \beta_1 X_1$ (a lin-log model)
d. $Y_1 = \beta_0 + \beta_1 \ln X_1$ (a log-lin model)
e. $\ln Y_1 = \beta_0 + \beta_1 \ln X_1$ (a double-log model)

2. Using the same variables as in question B.1, estimate a 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. 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 of *Writing about Multivariate Analysis*.
   c. Calculate differences in predicted values of $Y_1$ for one-unit increases in $X_1$.
   d. Consider whether other increments of $X_1$ are better suited to your research question and data (see pages 177–182 of *Writing about Multivariate Analysis*). If so, repeat part c using those increments instead of one-unit increases.
   e. 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 c or d.
   f. Use the model goodness of fit statistics to test whether the quadratic specification of $X_1$ statistically significantly improves the fit of the model compared to a linear specification of $X_1$. Contrast this against your conclusions based on the test statistic for $\beta_2$.
   g. Optional: Use a spreadsheet to perform questions B.2a through B.2d, following the instructions in appendix D of *Writing about Multivariate Analysis*.

3. Using data on the dependent variable used in the preceding question and a binary independent variable (denoted *DUMMY* in the equations below, coded 1 for a specified value and 0 for the reference category),
   a. Estimate an OLS model of the specification: $Y_1 = \beta_0 + \beta_1 DUMMY$.
   b. Write a sentence interpreting $\beta_1$, following the guidelines in chapter 9 of *Writing about Multivariate Analysis*.
   c. Using the estimated coefficients from part a, calculate predicted values of $Y_1$ for cases in the reference category and the other category of *DUMMY*. Compare these against the mean value of $Y_1$ for each of those categories of *DUMMY* from a bivariate calculation.

4. Using the same variables that you used for $Y_1$, $X_1$, and *DUMMY* in questions B.1 and B.3, estimate an OLS model with an interaction between $X_1$ and *DUMMY*.
   a. Write an equation to convey the model specification, including both main effects and interaction terms.
   b. Calculate predicted values of $Y_1$ for cases in the reference category and those in the other category of *DUMMY* across the observed range of $X_1$ in your data.

c. Create a chart showing the shape of the estimated relationship among $Y_1$, $X_1$, and *DUMMY,* using the results from part b. (See chapter 6 of *Writing about Multivariate Analysis* for guidelines.)
d. Calculate differences in predicted values of $Y_1$ for one-unit increases in $X_1$ for cases with each value of DUMMY.
e. Using the GEE approach, write a paragraph describing the interaction between $Y_1$, $X_1$,and *DUMMY,* using the guidelines under "Interactions" on pages 314–315 of *Writing about Multivariate Analysis*.
f. Optional: Use a spreadsheet to perform questions B.4b through B.4d, following the instructions in appendix D of *Writing about Multivariate Analysis*.

5. Using the same variables for $Y_1$ and *DUMMY* as in question B.4, and a three-category independent variable (*CATEGVAR*) from your data set, estimate an OLS model with an interaction between *DUMMY* and *CATEGVAR*. (Hint: Before you specify the model, create dummy variables for two of the three categories of *CATEGVAR,* following the guidelines on page 212 of *Writing about Multivariate Analysis*.)
   a. Write an equation to convey the model specification, including both main effects and interaction terms. Use this equation to help you define appropriate dummy variables to specify the interaction.
   b. Calculate the predicted values of $Y_1$ for all possible combinations of the variables *DUMMY* and *CATEGVAR*.
   c. Create a chart showing the shape of the estimated relationship between $Y_1$, *DUMMY,* and *CATEGVAR,* using the results from part b. See chapter 6 of *Writing about Multivariate Analysis* for guidelines.
   d. Using the GEE approach, write a paragraph describing the interaction between $Y_1$, *DUMMY* and *CATEGVAR*. See "Interactions" on pages 314–315 of *Writing about Multivariate Analysis* for example wording.
   e. Optional: Use a spreadsheet to perform questions B.5b and B.5e. See appendix D of *Writing about Multivariate Analysis* for spreadsheet guidelines.

6. Using data from your data set on a dichotomous dependent variable ($Y_2$), a continuous independent variable ($X_1$), and a categorical independent variable (*DUMMY*), estimate a logistic regression model of the form: $\text{logit}(Y_2) = \beta_0 + \beta_1 X_1 + \beta_2 DUMMY$. See your software manual for instructions on how to specify which category of your dependent variable to model.
   a. Write a sentence interpreting the value of $\beta_1$ using the guidelines on pages 221 and 226 of *Writing about Multivariate Analysis* for writing about odds ratios.
   b. Write a sentence interpreting the value of $\beta_2$.

## C. REVISING

1. Repeat questions A.1 through A.3 for a results section you have written previously that describes results from an OLS regression.

2. Repeat question A.6 for a results section you have written previously that describes results from a logistic regression analysis of a binary dependent variable.