10. *The "Goldilocks Problem" in Multivariate Regression*

PROBLEM SET

1. State whether a one-unit increase would be a useful contrast for each of the following topics. If not, suggest a more reasonable increment.

    a. Annual income (in dollars) for a family of four in the United States in 2004
    b. A Likert scale measuring extent of agreement with a gun control law
    c. Cholesterol level in milligrams per deciliter (mg/dL)
    d. Proportionate increase in the unemployment rate
    e. Hourly minimum wage (in dollars) in the United States in 2004

2. Zimmerman (2003) reports that the mean combined (verbal + math) SAT score for Williams College students in the classes of 1990–2001 was 1,396 points, with a standard deviation of 123. He estimates an OLS regression model of college GPA, with combined SAT score as an independent variable, with the results shown in table 9A. For each of the following situations, select pairs of plausible values of combined SAT scores to use as inputs for an illustration of effect size. Explain your reasoning, keeping in mind that each SAT score (math and verbal) can range from 200 to 800 points in increments of 10.

    a. A sample of students from an elite liberal arts college.
    b. A sample of all high school students nationwide who entered college.

Laditka et al. conducted a multivariate analysis of factors associated with the ambulatory care sensitive hospitalization rate in urban counties in the United States. Answer questions 3 through 8 using the guidelines in chapters 9 and 10 of *The Chicago Guide to Writing about Multivariate Analysis, 2nd Edition*.

TABLE 10A. Means and standard deviations of variables used in models predicting rate of ambulatory care sensitive hospitalization in US urban counties, 2000

Variable	Mean	Standard deviation
Outcome variables		
County-level ambulatory care sensitive hospitalization rate (ACSH) per 100,000 population, by age group (years)		
Ages 18–39	7.11	3.02
Ages 40–64	20.45	8.54
County-level health system and use factors		
Number of primary care MDs per 100,000 population	71.12	40.19
Number of short-term general hospital beds per 1,000 population	2.75	2.00
Percentage of hospitals that are investor owned	9.10	23.13
Medicaid generosity[a]	1.31	0.30
Number of community health centers	0.43	0.50
Number of emergency department visits per 1,000 population	381.51	177.01

[a] $1,000s of Medicaid expenditures per person under age 65 years below 200% of the poverty threshold.

Adapted from James N. Laditka, Sarah B. Laditka, and Janice C. Probst,. "More May Be Better: Evidence of a Negative Relationship between Physician Supply and Hospitalization for Ambulatory Care Sensitive Conditions," *Health Services Research* 40, no. 4 (2005): 1148–66, tables 2 and 3.

3. Answer the following questions based on table 10A from Laditka et al. (2005):

    a. What is the unit of analysis in this study?
    b. For each of the following variables, report the requested mean value and explain how you calculated it from the information in the table. Hint: What transformation was needed to get from the scale shown in the table to the scale requested in this question?
        i. Primary care MD's per person
        ii. Short-term general hospital beds per person
        iii. Medicaid generosity in dollars
        iv. Emergency room visits per person
    c. With reference to your answers to part b and the concepts covered in chapter 10 of *The Chicago Guide to Writing about Multivariate Analysis, 2nd Edition*, explain why you think the authors changed the scales of those variables for their analysis.

4. Calculate the value of the ambulatory care sensitive hospitalization rate (ACSH) one standard deviation below the mean and one standard deviation above the mean for

    a. Persons aged 18 to 39 years
    b. Persons aged 40 to 64 years

TABLE 10B. Standardized coefficients[a] from an OLS regression predicting rates of hospitalization for ambulatory care sensitive conditions in US urban counties, 2000

Variable	Standardized coefficients	
	Ages 18–39 years	Ages 40–64 years
County-level health system and use factors		
Number of primary care MDs per 100,000 population	−0.164***	−0.196***
Number of short-term general hospital beds per 1,000 population	0.227***	0.183***
Percentage of hospitals that are investor owned	0.083**	0.072*
Medicaid generosity[b]	−0.066†	−0.064†
Number of community health centers	0.044	0.037
Number of emergency department visits per 1,000 population	0.059	0.056
R^2	0.53	0.62

[a] Model also controls for county racial composition, age composition, crime rate, population density, population growth rate, household composition, household income, disability rate, death rates from heart disease, chronic obstructive pulmonary disease (COPD), diabetes, and liver disease, and for percentage of days with unhealthy air quality.

[b] $1,000s per person under age 65 years below 200% of the poverty threshold.

† $p < 0.05$; * $p < 0.01$; ** $p < 0.001$; *** $p < 0.0001$

Adapted from James N. Laditka, Sarah B. Laditka, and Janice C. Probst,. "More May Be Better: Evidence of a Negative Relationship between Physician Supply and Hospitalization for Ambulatory Care Sensitive Conditions," *Health Services Research* 40, no. 4 (2005): 1148–66, table 4.

5. Write sentences interpreting each of the following coefficients from the model for persons aged 18–39 shown in table 10B. Be sure to specify direction, magnitude, statistical significance, and units for both independent and dependent variables *as specified in the model*:

   a. Community health centers
   b. General hospital beds
   c. Primary care MD physicians
   d. Which variable had the largest effect per standard deviation unit increase?

6. Rewrite each of the sentences from the preceding question, rephrasing the results in the original units (not standardized units) of the dependent variable. Hint: Use the information in table 10A above.

7. Suppose Congress passed a law to add one community health center to every urban county. Write a sentence to predict the effect of that change on the ambulatory care sensitive hospitalization rate holding all other variables constant. Hint: Refer to table 10A to relate a one-unit increase to standard deviations of that independent variable.

8. Write a sentence reporting the effect on the ambulatory care sensitive hospitalization rate of moving from 2.75 to 4.75 short-term general hospital beds per 1,000 county residents.

Xu et al. (2006) analyzed the role of cohabitation in remarriage in the United States in the 1980s. Answer questions 9 and 10 based on the information in table 10C and the guidelines in chapters 9 and 10 of *The Chicago Guide to Writing about Multivariate Analysis, 2nd Edition*. Hint: Check the form of the dependent variable in the model.

TABLE 10C. Ordinary least squares regression coefficients for a model of waiting time to remarry (years, logged), United States, 1980s

Variable	Estimated coefficient
Intercept	1.843***
Respondent's overall cohabitation	
No cohabitation	
Cohabited prior to first marriage	−0.109
Cohabited prior to remarriage	−0.214***
Cohabitated prior to both marriages	0.028
Respondent's marital history	
Duration of first marriage (years)	−0.017†
Age at first divorce (years)	−0.019*
Residential children at time of divorce	
None	
Minor	−0.035
Adolescent	−0.130
Adult	−0.049

Adapted from Xiaohe Xu, Clark D. Hudspeth, and John P. Bartkowski.. "The Role of Cohabitation in Remarriage," *Journal of Marriage and Family* 68, no. (2006): 261–74, table 2.

Model also controls for gender, race/ethnicity, religious affiliation, employment status, educational attainment, and birth cohort. *F-statistic* = 6.342***; $R^2 = 0.079$; $N = 1,583$.
† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Mean log(waiting time to remarry, years) = 0.901; standard deviation = 1.043

9. Write sentences to interpret each of the following coefficients from table 10C:

   a. Cohabitation prior to remarriage
   b. Duration of marriage
   c. Presence of minor children at the time of first divorce

10. What is the predicted waiting time in years for respondents who did not cohabit prior to either their first marriage or remarriage, had no residential children at the time of divorce, and who were divorced at age 30 after being married for 5 years?