9

Quantitative Comparisons for Multivariate Models

PROBLEM SET

1. Indicate whether each of the following statements is correct. If not, rewrite the second part of the sentence to agree with the first.
   a. "The odds ratio of passing the test was 0.60 for students in School A compared to School B, meaning that students in School A were 60% more likely to pass than those in School B."
   b. "Log-odds of migration for men whose siblings had migrated were 0.51, reflecting higher chances of migration for them than for men whose siblings had not migrated."
   c. "Relative odds of migration for ever-married men were 0.91, reflecting higher chances of migration for ever-married than never-married men."
   d. "The standardized beta for widows was −0.5, meaning that widows scored on average half a point lower than nonwidows."
   e. "The relative risk of divorce for teens compared to older adults was 2.50, corresponding to an excess risk of 150% for teens."
   f. "The relative risk dropped from 2.50 to 2.00 between the unadjusted and adjusted models, corresponding to a 50% reduction in excess risk."

2. For each of the following research questions, indicate whether you would specify an OLS model or a logit model, and identify the units or omitted category of the dependent variable.
   a. Whether income is associated with chances of being arrested.
   b. Whether a new medication decreases average cholesterol levels.
   c. Whether child's IQ varies by parents' IQs.
   d. Whether cohabitation prior to marriage is associated with risk of divorce.

Table 9A. Regression of cumulative grade point average by own SAT scores and roommate's SAT scores, Williams College classes of 1999–2001

	Coeff. (s.e.)
Own verbal SAT score/100	0.195 (0.011)
Own math SAT score/100	0.092 (0.011)
Race (ref. = white)	
Black	–0.264 (0.033)
Hispanic	–0.160 (0.035)
Native American	0.098 (0.175)
Not a U.S. citizen	0.099 (0.043)
Asian	–0.085 (0.022)
Female	0.128 (0.013)
Roommate's verbal SAT score/100	0.027 (0.010)
Roommate's math SAT score/100	–0.016 (0.010)
Sample size	3,151
R^2	0.378

Source: Adapted from David A. Zimmerman, "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics* 85.1 (2003): 9–23, table 3. Also available to subscribers at *http://weblinks2.epnet.com*.

In a 2003 article in the journal *Review of Economics and Statistics,* Zimmerman uses data from Williams College on individual students' grades, their SAT scores, and their roommates' SAT scores to estimate models of peer effects on academic performance (table 9A). Use that information to answer questions 3 through 7 below.

3. For the model shown in table 9A,
   a. Identify the dependent variable, the type of variable (continuous or categorical), its units or coding, and theoretically possible range.
   b. State whether an OLS model or logit model is more suitable for this analysis.

c. Identify the continuous independent variables, their units as specified in the model, and their theoretically possible ranges.
d. Identify the categorical independent variables and their reference categories.

4. What is the estimated difference between male and female GPAs? Is that difference statistically significant?

5. What is the difference in predicted GPAs if a student's own verbal SAT score was 720 instead of 680? (Assume the student is in the reference category for all other variables in the model.)

6. What is the difference in predicted GPAs if a student's roommate's math SAT score was 720 instead of 680? (Assume the student is in the reference category for all other variables in the model.)

7. If the intercept term is 0.780, what would the predicted GPA be for a white male student with a verbal SAT of 720, a math SAT of 700, and a roommate with a verbal SAT of 680 and a math SAT of 650? (Actual intercept terms could not be reported due to confidentiality of students' information.)

Using data from the 1979 National Longitudinal Survey of Youth, Light (2004) analyzes gender differences in effects of marriage and cohabitation on change in total family income (table 9B). Answer questions 8 and 9 using that information.

Table 9B. Estimated effect of marital-status transitions on total family income, United States, 1979–2000

Gender and type of marital status transition	Coefficient	Standard error
Women		
Single to cohabiting	0.440	0.027
Single to married	0.416	0.026
Men		
Single to cohabiting	–0.011	0.026
Single to married	–0.035	0.025
Women and Men		
Cohabiting to married	–0.013	0.019

Source: Adapted from "Difference Model 2," table 3, from Audrey Light, "Gender Differences in the Marriage and Cohabitation Income Premium," *Demography* 41.2 (2004): 263–84, *http://muse.jhu.edu/journals/demography/v041/41.2light.pdf.*

Notes: $N = 4{,}700$ women and 5,139 men. Dependent variable = log(posttransition income) – log(pretransition income).

8. Perform these tasks using the information in table 9B.
a. Write a sentence identifying the dependent variable in the model without using an equation.
b. Calculate the value of the dependent variable corresponding to an increase in income from $20,000 to $35,000.

9. Write sentences to present the effects of the following transitions, using the information in table 9B.
   a. The effect of a woman transitioning from single to married on change in log(family income).
   b. The effect of a woman transitioning from single to married, in terms of percentage change in family income.
   c. The effect of a man transitioning from single to married, in terms of percentage change in family income.
   d. The effect of a woman transitioning from single to married on family income in dollars, assuming that she had an income of $20,000 when she was single.

Fussell and Massey (2004) used data from the Mexican Migration Project to study relationships among demographic factors, human capital, social capital in the family and community, and migration from Mexico to the United States (table 9C). Use the information in table 9C to answer questions 10 through 13.

Table 9C. Estimated log-odds of first trip to the United States, Men, 1987–1998 Mexican Migration Project

	Log-odds	Standard error
Demographic background		
Age (years)	–0.003	0.02
Age-squared	–0.001	0.0002
Ever married	–0.09	0.06
Number of minor children in household	0.01	0.01
Human capital		
Years of education	–0.04	0.006
Months of labor-force experience	–0.002	0.0007
Social capital in the family		
Parent a prior U.S. migrant	0.51	0.05
Siblings prior U.S. migrants	0.36	0.02
Social capital in the community		
Migration prevalence ratio[a]		
0–4	–0.99	0.15
5–9	–0.09	0.12
(10–14)		
15–19	0.35	0.10
20–29	0.57	0.13
30–39	0.95	0.15
40–59	0.74	0.19
60 or more	0.34	0.15
Intercept	–3.31	0.26
– 2 log likelihood	23,369.2	
Df	26	

Source: Adapted from Elizabeth Fussell and Douglas S. Massey, "The Limits to Cumulative Causation: International Migration from Mexican Urban Areas," *Demography* 41.1 (2004): 151–71. Table 2, http://muse.jhu.edu/journals/demography/v041/41.1fussell.pdf.

Note: Model also includes controls for occupational sector, internal migratory experience, community characteristics, and Mexican economic and U.S. policy context.

[a] The migration prevalence ratio = (the number of people aged 15+ years who had ever been to the U.S./the number of people aged 15+ years) ×100.

10. Perform these tasks using the information in table 9C.
    a. Identify the dependent variable, the type of variable (continuous or categorical), its units or coding, and theoretically possible range.
    b. State whether an OLS model or logit model is more suitable for this analysis.
    c. Identify the continuous independent variables, their units as specified in the model, and their theoretically possible ranges.
    d. Identify the categorical independent variables and their reference categories.

11. Assuming all other variables are in the reference category or at their mean values, calculate the relative odds of first migration to the United States for
    a. an ever-married man compared to a never-married man
    b. a 30-year-old man compared to a 20-year-old man
    c. a man with a parent who is a prior U.S. migrant compared to a man without parents who migrated there
    d. a man from a community with a migration prevalence ratio (MPR) of 0–4 compared to one from a community with an MPR of 10–14
    e. a man from a community with a migration prevalence ratio (MPR) of 0–4 compared to one from a community with an MPR of 60 or more

12. Create a table contrasting odds of first trip to the United States at 10-year age intervals from 15 through 64 years; specify the values of the other variables you used in your calculations.

13. Calculate the odds of first migration for a 20-year-old never-married man with no children, eight years of education, 24 months of labor force participation, neither parents nor sibling prior migrants, from a community with an MPR of 10–14.

Table 9D summarizes results of Carr's (2004) analysis of relations among dependence on a spouse, gender, and psychological adjustment to the death of a spouse. Answer questions 14 and 15 using that information.

Table 9D. OLS regressions of self-esteem at wave 2, overall and by gender, changing lives of older couples (cloc) study, 1987–1994

	Total sample		Women		Men	
Variable	Coeff.	Std. error	Coeff.	Std. error	Coeff.	Std. error
Widow	–0.51*	0.24	0.25†	0.15	1.67	1.22
Female	–0.60**	0.22				
Interaction: female × widow	0.70**	0.26				
Emotional dependence on spouse			–0.35**	0.13		
Interaction: emotional dependence on spouse × widow			0.34**	0.15		
Dependence on spouse for homemaking tasks					2.67*	1.35
Interaction: dependence on spouse for home making tasks × widow					–2.92*	1.39
Dependence on spouse for home maintenance and financial tasks					–1.30*	0.55
Interaction: dependence on spouse for home maintenance and financial tasks × widow					1.58**	0.59
Intercept	2.13	0.76*	0.54	0.79	1.75	2.12
R^2 adjusted	0.19		.024		0.19	
Unweighted *N*	297		217		80	

Source: Adapted from Deborah Carr, "Gender, Preloss Marital Dependence, and Older Adults' Adjustment to Widowhood," *Journal of Marriage and the Family* 66 (2004): 220–35, table 2. Models also control for wave 1 well-being, demographic characteristics, and number of months between wave 1 and 2 interviews. Dependence measures assessed at wave 1.

* $p < 0.05$. ** $p < 0.01$. † $p < 0.10$.

14. Using the results for women in table 9D:
    a. Create a spreadsheet to calculate the net effect of the interaction between emotional dependence on spouse, widowhood status, and predicted self-esteem, using the guidelines in appendix D of *Writing about Multivariate Analysis*. Both self-esteem and emotional dependence are in standardized units (mean = 0, standard deviation [SD] = 1). Allow emotional dependence to vary from –1.0 to 1.0 SD in your calculations.
    b. Design a chart to portray this pattern.
    c. Write a short description of the association between emotional dependence on spouse, widowhood status, and predicted self-esteem using the GEE approach.
    d. Explain why there isn't a dummy variable for "female" in the stratified models.

15. Using the results for the total sample in table 9D:
    a. Create a table to show predicted self-esteem for each of the four possible combinations of gender and widowhood status.
    b. Create a chart to portray that association.
    c. Write a short description of the association between gender, widowhood status, and predicted self-esteem using the GEE approach.

16. Suppose a study found that the unadjusted odds ratio of hospital admission for diabetics compared to nondiabetics is 3.50.
    a. Calculate the excess risk of hospital admission for diabetics.
    b. When demographic factors and other health conditions are taken into account, the adjusted odds ratio for diabetics is 3.00. Calculate the change in excess risk of hospital admission for diabetics between the adjusted and unadjusted models.

17. Suppose a study found that 20% of nondiabetics were admitted to the hospital.
    a. Using the adjusted odds ratio from the previous question, calculate the corresponding relative risk of hospital admission for diabetics.
    b. Express the discrepancy between the odds ratio and the relative risk as a percentage difference.
    c. Write a sentence describing the association between diabetes and hospital admission, using the criteria under "An Aside on Relative Risk and Relative Odds " on pages 224–226 of *Writing about Multivariate Analysis*.