

15. Writing about Multivariate Models

PROBLEM SET

Fauth et al. (2004) studied the effects of a residential mobility experiment, comparing outcomes of low-income adults in public housing who moved to low-poverty neighborhoods to those who stayed in their original, high-poverty neighborhoods. “Movers” were chosen by lottery from among those who applied for the program. Their results are summarized in tables 15A (bivariate statistics) and 15B (multivariate model results). Use those data to answer questions 1 through 3.

TABLE 15A. Individual background characteristics, neighborhood, and housing characteristics of movers and stayers, Yonkers Residential Mobility Program, 1994–1995

	Residential status			χ^2 or F^a
	Movers ($n = 173$)	Stayers ($n = 142$)	Total ($n = 315$)	
<i>Background characteristics</i>				
Age (mean years)	36.69	34.07	35.51	6.45**
Female	97%	96%	97%	0.41
Latino (ref. = black)	31%	25%	28%	1.07
At least high school education	67%	53%	61%	6.62**
Female household head	76%	85%	80%	4.39*
Mean # children in household	1.72	2.01	1.85	6.04*
<i>Neighborhood/housing^b</i>				
Danger	0.26	1.29	0.72	144.11***
# of victimizations in past year	0.12	0.32	0.21	9.21*
Disorder	0.15	1.41	0.72	796.17***
Cohesion	0.62	0.40	0.52	43.48***
Resources	3.05	2.89	2.98	4.90*
Housing problems ^c	0.20	0.54	0.35	54.40***

Source: Adapted from Rebecca C. Fauth, Tama Leventhal, and Jeanne Brooks-Gunn, “Short-Term Effects of Moving from Public Housing in Poor to Middle-class Neighborhoods on Low-Income, Minority Adults’ Outcomes,” *Social Science and Medicine* 59 (2004): 2271–84, table 1. <http://www.sciencedirect.com/science>.

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

^a χ^2 statistic reported for difference in categorical variable between movers and stayers; F -statistic for difference in continuous variable.

^b Ranges of values for the neighborhood and housing quality measures are as follows: Danger: 0 to 3; disorder: 0 to 5; cohesion: 0 to 4; resources: 0 to 5; housing problems: 0 to 5.

^c In the published paper, this measure was termed “housing quality,” but I relabeled it “housing problems” to reduce confusion because a higher value indicates more problems, e.g., with rats and mice.

1. Answer the following questions based on the information in table 15A:
 - a. Did the random assignment succeed in equalizing the background characteristics of movers and stayers? Write a paragraph summarizing the similarities and differences in background characteristics between those two groups.
 - b. Did neighborhood and housing characteristics differ according to residential status (e.g., for movers versus stayers)? Write a paragraph generalizing these findings.
 - c. What do these statistics suggest about the need for multivariate models of these outcomes by residential status? Explain your reasoning.
2. Write a paragraph describing the results in table 15A, using your answers to question 1 and the principles on p. 312 of *Writing about Multivariate Analysis, 2nd Edition* for building the case for a multivariate model.
3. Write a description of the findings in table 15B, using the GEE approach to summarize findings across the six dependent variables, following the guidelines in chapters 2, 14, and 15 and appendix A.

TABLE 15B. Results from OLS models of six neighborhood characteristics and housing problems measures, Yonkers Residential Mobility Program, 1994–1995

Independent variable	Dependent variable					Housing problems ^a
	Danger	Victimization	Disorder	Cohesion	Resources	
Mover	−0.99***	−0.19**	−1.25***	0.21***	0.13	−0.30***
Age (years)	0.01	0.00	0.00	0.00	0.00	0.00
Latino	0.16	0.00	−0.02	−0.01	0.09	−0.19***
High school graduate	0.06	0.07	0.04	0.02	0.05	−0.06
Female headed HH	−0.27*	−0.01	0.02	−0.03	−0.05	0.07
# children in HH	0.05	0.07*	0.05*	−0.01	0.00	0.03
R^2	0.34	0.05	0.73	0.14	0.02	0.20

Source: Adapted from Fauth, Leventhal, and Brooks-Gunn 2004, table 3.

^a In the published paper, this measure was termed “housing quality,” but I relabeled it “housing problems” to reduce confusion because higher value indicates more problems, e.g., with rats and mice.

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

4. Write a description of Zimmerman’s findings (table 15C), focusing on the results for own SAT scores and roommate’s SAT scores. Follow the guidelines in chapter 15 about organizing your description. Generalize across the three models to the extent possible: Which results are similar for the three groups, and which differ? Why did Zimmerman run three models?

TABLE 15C. Estimated coefficients and standard errors from a model of cumulative grade point average by own SAT scores and roommate's SAT scores, stratified by own SAT score, Williams College classes of 1999–2001

	Student's own combined math & verbal SAT score		
	Lowest 15%	Middle 70%	Top 15%
Own verbal SAT score/100	0.205 (0.039)	0.199 (0.015)	0.118 (0.055)
Own math SAT score/100	0.065 (0.036)	0.112 (0.017)	0.045 (0.051)
Race (ref. = white)			
Black	−0.181 (0.046)	−0.386 (0.053)	−0.800 (0.059)
Hispanic	−0.036 (0.059)	−0.254 (0.046)	−0.050 (0.274)
Native American	−0.238 (0.169)	0.212 (0.168)	dropped
Not a US citizen	0.076 (0.091)	0.126 (0.055)	0.055 (0.066)
Asian	0.210 (0.120)	−0.065 (0.026)	−0.201 (0.047)
Female	0.262 (0.038)	0.103 (0.016)	0.107 (0.028)
Roommate's verbal SAT score/100	0.006 (0.025)	0.043 (0.012)	−0.013 (0.021)
Roommate's math SAT score/100	−0.038 (0.028)	−0.021 (0.012)	0.030 (0.022)
Sample size	450	2,072	629
R^2	0.41	0.27	0.21

Source: Adapted from David A. Zimmerman, "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics* 85, no. 1 (2003): 9–23, table 4.

Answer questions 5 through 7 based on the results in table 15D from Fus-sell and Massey (2004).

TABLE 15D. Estimated log-odds of first trip to the United States, men, 1987–1998 Mexican Migration Project

	Log-odds	Standard error
<i>Demographic background</i>		
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
<i>Human capital</i>		
Years of education	−0.04	0.006
Months of labor-force experience	−0.002	0.0007
<i>Social capital in the family</i>		
Parent a prior US migrant	0.51	0.05
Siblings prior US migrants	0.36	0.02

	Log-odds	Standard error
<i>Social capital in the community</i>		
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, no. 1 (2004): 151–71, table 2. <http://muse.jhu.edu/journals/demography/v041/41.fussell.pdf>.

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

^aThe migration prevalence ratio = (the number of people aged 15+ years who had ever been to the US/the number of people aged 15+ years) × 100.

- Write a description of the age pattern of migration to the United States, with reference to the chart you created in question 9a of the problem set for chapter 6.
- Write a description of the relationship between human capital and migration.
- Write one to two paragraphs describing the association between social capital in the family and community and migration from Mexico to the United States, with reference to the results in table 15D and the chart you created in question 9b of the problem set for chapter 6.

Pan et al. (2005) estimated a series of multilevel growth trajectory models of toddler vocabulary. The model specification and goodness of fit statistics are shown in table 15E.

TABLE 15E. Model specification and goodness-of-fit statistics for four multilevel growth trajectory models of toddler vocabulary development among children from low-income families

Variables	Model 1	Model 2	Model 3	Model 4
Intercept	X	X	X	X
Age and age-squared	X	X	X	X
Mother tokens (3 variables)		X		
Mother word types (3 variables)			X	
Mother points (3 variables)				X
Random effects parameters (4 variables)	X	X	X	X
–2 Log likelihood	1,932.9	1,931.6	1,928.4	1,929.1
Akaike Information Criterion (AIC)	1,952.9	1,957.6	1,954.4	1,955.1
Degrees of freedom	7	10	10	10

Adapted from Barbara Alexander Pan, Meredith L. Rowe, Judith D. Singer, and Catherine E. Snow, “Maternal Correlates of Growth in Toddler Vocabulary Production in Low-Income Families,” *Child Development* 76, no. 4 (2005): 763–82, table 2.

8. Use the information in table 15E to answer the following questions:

- a. Which models are nested? Explain why.
- b. Which models are not nested? Explain why.
- c. Keeping in mind your answers to parts a and b, identify the parsimonious model among fit of models 1 through 4 using the guidelines on “Comparing Models using *AIC* or *BIC*” from p. 335 of *Writing about Multivariate Analysis, 2nd Edition*.