Teaching How to Write about Multivariate Analysis: Suggested Courses and Exercises

Jane E. Miller[1] and Lawrence Wu[2]

[1] Professor, Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson Street, New Brunswick NJ 08901; (848) 932-6730; fax: (848) 932-1253; jmiller@ifh.rutgers.edu.

[2] Professor, Department of Sociology, New York University, 295 Lafayette St., Room 4103, New York, NY 10012; (212) 992-9565; lawrence.wu@nyu.edu.

Acknowledgements: This paper is based on an invited panel academic workshop that was presented at the 2008 Annual Meeting of the American Sociological Association. We thank Paula England and Donald Treiman for helpful comments on earlier drafts.

ABSTRACT

Writing effectively about multivariate analysis involves a range of skills, including those that are typically taught in graduate courses or seminars about expository writing, multivariate regression, and research methods. This paper discusses how to integrate teaching of these skills into a graduate sociology program, including suggestions for courses and informal teaching settings into which the material can be incorporated, tested classroom teaching approaches, and exercises to provide practice applying those skills to course material or ongoing research projects.

INTRODUCTION

Writing effectively about multivariate analysis involves a range of skills, including those that are typically taught in graduate courses or seminars about expository writing, multivariate regression, and research methods. In a related piece (Miller et al., 2009), we outlined a series of important issues to consider when writing papers involving a multivariate regression analysis, and provided a series of guidelines and examples for avoiding common pitfalls in that type of writing. This paper discusses how to integrate teaching of these skills into a graduate sociology program, including suggestions for courses and informal teaching settings into which the material can be incorporated, tested classroom teaching approaches, and exercises to provide practice applying those skills to course material or ongoing research projects.

WRITING ABOUT MULTIVARIATE ANALYSIS IN A GRADUATE CURRICULUM

Writing about multivariate analysis can be integrated into several types of courses and informal learning settings as part of a graduate curriculum.

Formal graduate courses

- In a *multivariate regression course*, assign students short, self-contained exercises on how to report and interpret coefficients in simple sentence form. These tasks can be added to existing data analysis assignments. These exercises help get students in the habit of thinking about how to communicate clearly about multivariate results, and can also help reinforce the substantive meaning of the statistical concepts.
- In *research methods courses*, ask students to identify the purpose of a multivariate model for different topics and data (e.g., to test for confounding or mediating). These exercises help students reinforce their understanding of the reason for a particular model specification in the context of a given research question and data set.
- In both *research methods courses* and *substantive courses* such as sociology of the family, medical sociology, or criminology, ask students to evaluate and critique articles that use multivariate regression to analyze data from different types of study designs, assessing whether causal arguments are plausible given the data and methods at hand.

- In *substantive courses*, require students to write either the pre-results or results/conclusion part of a paper, with multiple drafts expected. This type of assignment forces students to think though how methodological and statistical issues affect the analytic plan designed to answer an underlying substantive question.
- In *writing courses*, assign students to decide among tables, graphs, and prose for specific tasks related to a paper they are writing about an application of multivariate analysis, and then to draft those materials according to guidelines provided. These tasks help students understand the strengths and weaknesses of each of the tools for presenting numbers, and to become proficient at designing effective versions of graphs and tables and writing the associated prose.
- In *advanced graduate courses* such as professional development seminars, ask students to revisit these concepts and skills while writing or revising their own papers, to provide practice identifying and explaining the big picture of how a multivariate model fits their topic and data. Such courses permit time to consider how writing about multivariate analysis fits into each section of a research paper, and an opportunity to apply standard expository writing techniques to this specific type of academic writing.

Informal settings

Informal teaching settings can also provide opportunities for teaching and reinforcing issues related to effective design, execution, and communication of multivariate analyses applied to sociological research questions. Such settings include brown bag or other research seminars, collaborative research involving graduate students on teams with faculty or post-doctoral or other graduate students, and individual meetings with faculty supervising course- or qualifying papers or doctoral dissertations.

Seminars. Brown bag seminars provide opportunities for students to hear how experienced researchers plan, conduct, communicate, and critique multivariate analyses. Aspects of research thinking and writing that are often discussed in these settings include:

- *Research methods issues* that affect multivariate analyses, such as whether the choice of data set, sampling strategy, and analytic sample were appropriate for the research question; and whether the variables used in the model are adequate measures of the underlying concepts.
- *Statistical issues* such as suitability of the type of model to the variables involved; whether key variables were omitted from the analysis; and the choice of analytic strategy to address the particular research question (e.g., nested models to test for mediating effects).
- *Substantive issues* such as whether the research question is interesting and important to sociologists; whether results are credible (and if not, why not); whether there are alternative explanations for the findings; and whether the limitations of the data were taken into account in the interpretation of results.
- *Communications issues* such as clarity of graphs, tables, and prose used to convey the purpose, results and implications of the study; how experienced researchers organize tables to convey their analytic plan and statistical findings; and how they read (and interpret) the information in those tables to answer the underlying research question.

Early in each academic year, it is useful to orient graduate students about how to listen to and participate in seminar experiences so that they don't simply focus on the substantive conclusions. Point out the importance of also attending to the kinds of issues raised by members of the audience, how the presenter answers those questions; and how the feedback received by the presenter might affect his or her work on the topic. Encourage students to develop and ask questions during the seminar. Suggest that they think about which aspects of the presentation were effective (e.g., clear explanation of why a multivariate model for this topic and data) and which perhaps should be avoided in their own future written and oral research presentations (e.g., too many digits on the slides; overly technical descriptions of the results that obscure substantive meaning).

Collaborative research settings. Meetings and written feedback on drafts of multivariate research projects by a team of researchers can also provide insight into many of the issues outlined above, while bringing the advantage of the student being an active participant in that research who is familiar with the topic and data. Hearing how a group of collaborators each contributes their own perspective, how they complement one another's strengths, critique one another, and resolve conflicts about issues that arise during the project also provide valuable opportunities for students to learn about the wide range of tasks and skills involved in writing about multivariate research.

Individual research projects. Work on individual research papers or dissertations also allows graduate students to receive feedback on their thinking process, analytic approach, interpretation, and writing related to applications of multivariate analysis to sociological research questions. The individualized nature of these meetings is invaluable for ensuring that each student receives guidance in the particular issues they are facing in their own projects at the moment and how the principles for writing about multivariate analysis apply to their specific project. Such feedback is critical to fostering their development into researchers who are capable of conducting well-conceived and executed independent research projects, and can anticipate and address the issues that are central to multivariate research projects.

METHODS FOR TEACHING HOW TO WRITE ABOUT MULTIVARIATE ANALYSIS

There are several steps to teaching how to write about multivariate analysis in graduate coursework or for dissertation writers. First, assign readings that cover key principles about statistical research writing, such as Miller (2005), Treiman (2009), or other books or articles on writing or professional research practice. Second, in lecture, briefly cover the principles and associated skills for writing about multivariate analysis, followed by in-class demonstration using such as the "poor/better/best" technique (shown below) to show students examples of how to translate abstract writing principles into concrete sentences or paragraphs; see Miller (2005) or Miller, England, Treiman and Wu (2009). Third, reinforce those concepts by assigning students to apply them to their own work or to evaluating existing published work, using one of several types of exercises, such as those shown below. Fourth, have the students use checklists such as those at the end of each chapter in *(*Miller, 2005) to plan and evaluate their work.

Poor/better/best in the classroom

The "poor/better/best" ("PBB") approach can be used to give practice applying new principles about how to write about some aspect of a multivariate analysis, encouraging students to draft, evaluate, and revise sentences, tables or charts as a participatory exercise. Here is a suggested sequence of steps for applying poor/better/best in a classroom setting.

1. Introduce a general principle such as generalizing a pattern based on many numbers. Students should have been assigned a reading on the topic, in this case, (Miller, 2005, Chapter 2).
2. Walk students through one complete PBB exercise, using examples from the article or book, or ones you have written based on an assigned substantive article. **HINT:** Start with bivariate, and then progress to one or more multivariate examples.

Sample poor/better/best teaching example for generalizing a pattern

> <u>*Poor*</u>: "In 1985, 14-17 year-olds committed 952 gun-related homicides. In 1986 and 1987, persons in that age group committed 1,099 and 1,207 gun-related homicides, respectively (Figure 1)… [***Description continues by reporting annual statistics for each of the three age groups***]."
>
> <u>*Comment*</u>*: Individual statistics on the number of homicides in each of fifteen years for each of several age groups force readers to do the math to figure out whether homicides are rising, falling, or level, and whether the time trend is similar for the three age groups being compared. All those sentences reporting numbers will also obscure the general pattern and tire your readers.*

Figure 1 about here

Poor [version 2]: "Between 1985 and 1986, the number of gun-related homicides committed by 14-17 year olds increased from 952 to 1,099. Between 1986 and 1987, it increased again, to 1,207… (Figure 1)… [***Description continues by reporting one-year changes in the number of homicides for each of the three age groups***.]"

Comment: Although this version presents single-year changes in homicides committed by one age group instead of merely reporting the value of each data point, it fails to paint the overall shape of the time trend across the period shown or to compare across age groups.

(Somewhat) better: "Among offenders aged 14-17, gun-related homicides nearly quadrupled between 1985 and 1994 (from 952 to 3,617 homicides), and then declined to 1,079 in 2000. Among offenders aged 18-24, gun-related homicides more than doubled between 1985 and 1994 (from 3,633 to 8,253), then decreased through 2000. Among offenders aged 25 and older, gun-related homicides declined slightly throughout the period from 1985 to 2000 (Figure 1)."

Comment: Although this version describes the shape and size of the time trend for each age group, it doesn't compare time trends across age groups, requiring readers to figure out for themselves whether all three age groups followed the general time trends observed in Figure 1 or whether the patterns varied.

Best: "As shown in Figure 1, in the two youngest groups of offenders, gun-related homicides increased substantially between 1985 and 1994, and then decreased steadily until 2000. In contrast, the number of gun-related homicides committed by older offenders decreased slowly throughout the time period shown."

Comment: This description points out that the time trend in gun-related homicides was similar for two of the three age groups, and then describes the general shape of pattern. The phrase "in contrast" is used to emphasize the fact that the time trend for the third age group was different from the other two, before going on to describe the shape of that pattern.

3. Hand out a table or chart with numeric information for an association, and then give students a few minutes to work independently (individually) to write a sentence to report and interpret those numbers. See "Problem Sets" below for examples of the types of tasks that can be used for this exercise.
4. Solicit an example sentence and write it on the board, then ask the class to use the mental checklist for that principle to evaluate the sentence and provide constructive criticism on how to improve it. See "Checklists" below.

   HINT: Try wording your request "Can someone give me a poor sentence to report that information?," which tends to make students more willing to hazard a suggested sentence than if they are asked to state their best attempt at an effective sentence.
5. Write the revisions in a contrasting color. Point out that revision is a normal part of the writing process and is to be expected especially when tasks such as interpreting coefficients or describing non-linear patterns from multivariate models are attempted for the first time.
6. Ask the students whether they feel prepared to apply the principle on their own. If not, have them identify the aspects of the principle that they don't understand, or the steps of the approach where they became stuck.
7. Have a second example table or chart at hand in case they want to practice again as a group.

Checklists

Checklists that summarize the key steps to achieving a particular task can be invaluable for helping students plan, draft, and evaluate their writing about multivariate analysis. For example, the checklist for summarizing a complex pattern reads:

- To describe a pattern involving many numbers, summarize the overall pattern rather than repeating all the numbers.

- Create a graph to illustrate the pattern. It doesn't have to be perfect but should be to scale and labeled so you can identify the variables involved and their categories and units.
- Find a generalization that fits most of the data, specifying direction and magnitude of the association.
- Report a few illustrative numbers from the associated table or graph;
- Describe exceptions to the general pattern – how some parts deviate from the shape and size of the pattern described in your generalization.

See Miller (2005) pages 29-32 and 311-315 for additional guidelines and examples on summarizing patterns.

EXERCISES FOR TEACHING HOW TO WRITE ABOUT MULTIVARIATE ANALYSIS

Problem sets

The first type of exercise involves use of problem sets – self-contained questions that provide practice in specific skills, working from data provided in text, table, or graph form. Solutions to these problems are given in the Appendix. Problem Set A might be used in an introductory course on multivariate regression.

Problem Set A: 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 1). Use that information to answer the following questions.

1. For the model shown in Table 1,
    a. Identify the dependent variable, the type of variable (continuous or categorical), its units or coding, and theoretically possible range.
    b. State whether an ordinary least squares (OLS) model or a logit model is more suitable for this analysis, and why.
    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.

Table 1 about here

2. Write sentences to present the direction, magnitude, and statistical significance of the following variables on cumulative college GPA, using the information in Table 1.
    a. Own verbal SAT score
    b. Female
    c. Black compared to white race.
    d. Black compared to Hispanic race/ethnicity.

Problem Set B could be assigned to students in a research methods course, to help them understand about randomized experiments and confounding.

Problem Set B: 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. Table 2 reports results of bivariate analyses among the variables used in their analysis. Use those data to answer the following questions.

1. What does "chosen by lottery" tell us about the design of this study?
2. What is the role of residential status in this analysis?
3. What is the role of the neighborhood/housing items in this analysis?
4. What is the role of the "background characteristics" in this analysis?
5. Write sentences to report the direction, magnitude, and statistical significance of the difference between movers and stayers according in terms of:

a. Age
b. Gender
c. Number of victimizations

6. Answer the following questions based on the information in Table 2:
   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.
   d. Suppose that the authors simply observed residents who moved and who stayed instead of the assignment of movers and stayers by lottery. Discuss the implications, if any, of these alternative designs.

Problem sets can be used in class immediately after teaching a skill or concept, with associated graphs or tables inserted onto slides for group discussion and/or given as handouts for students to write on as they answer the question. Other uses for problem sets include assigning them as homework or using them as exam questions. A series of problem sets related to Miller (2005) is available at http://www.press.uchicago.edu/books/miller/multivariate/, along with solutions to the odd numbered questions that can be used to assist in grading if those problems are assigned for homework or exams. These problem sets have been tested out in graduate-level courses for students in sociology and related fields. Another set, together with illustrative answers, is available at http://www.josseybass.com/go/quantitativedataanalysis (Treiman 2009).

Suggested course extensions

A second type of exercise is course extensions that ask students to apply the concepts and skills of multivariate analysis to the published literature or papers they are writing as part of courses in regression methods, research methods, research writing, or professional development seminars. These suggested course extensions come in four general types (1) reviewing published literature; (2) applying statistics using one's own data; (3) writing new material; and (4) revising previous drafts of material.

Reviewing published literature involves using the principles for writing effectively about multivariate analysis as part of a research methods course or substantive course in which students are asked to critique applications of multivariate analysis to important sociological research questions.

Example for a research methods course or a substantive sociology course: Find a journal article that uses multivariate regression to analyze a social policy problem and proposes one or more solutions to that problem.

a. Evaluate how well the article addresses each of these aspects of "importance." Does the article
   i. specify a cause-and-effect type of relationship?
   ii. provide a plausible argument for a causal association?
   iii. review measurements issues for the outcome and key covariates?
   iv. discuss bias, confounding, or reverse causation?
   v. report results of statistical tests for that association?
   vi. assess whether the expected benefits of the proposed solution are big enough to outweigh costs or otherwise matter in a larger social context?

b. Given your answers to part a, write a short critique of the appropriateness of the proposed solution.

Note that exercises like this one can be incorporated into substantive sociology courses on any of a wide range of topics, requiring students to integrate concepts from their research methods and statistics

courses along with knowledge about the specific content area covered in the course. Such an approach provides practice in the kinds of critical reading and analytic skills that sociologists must develop as the basis for future professional activities such as reviewing grant proposals and articles. It can also help to dispel the tendency of some graduate students to want to divorce methods material from substantive sociological material.

Applying statistics involves having students apply these planning and writing principles to results from their own data analyses as part of a course in regression methods.

> **Example for a statistics course that includes multiple regression:** Using data on a continuous dependent variable (denoted Y_1) and a continuous independent variable (denoted X_1), and a binary variable *DUMMY*, estimate an OLS model with an interaction between X_1 and *DUMMY*.
>
> a. a. Write an equation to convey the model specification, including both main effects and interaction terms.
> b. 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. c. Create a graph showing the shape of the estimated relationship among Y_1, X_1, and *DUMMY*, using the results from part b. (See Miller 2005:Ch. 6 for guidelines.)
> d. d. Write a paragraph describing the interaction between Y_1, X_1, and *DUMMY*.

Writing exercises ask students to apply the principles described above to plan and execute new sections of papers based on multivariate analysis of their own topic and data, using checklists such as those provided in (author citation).

> **Example for a multivariate regression course or advanced writing seminar:** Write an explanation of how you arrived at your final model specification, including the following topics:
>
> a. The criteria used to determine which variables were included in the model, with reference to your specific research question.
> b. Whether and why nonlinear specifications were used for any of the independent variables.
> c. Whether interactions were included among two or more independent variables, and if so, which ones and why.

Revising exercises require students to use those principles to evaluate and revise sentences, paragraphs, tables and graphs from their own previously written papers as part of a graduate writing course or professional development seminar. This type of activity works well as a peer-editing exercise in which student exchange initial drafts with one another, mark them up according to the principles described above, and then return them to the original author who revises the material accordingly.

> **Example for a multivariate regression course or advanced writing seminar:** Evaluate a description of a series of nested models from the results section of a paper you have written previously, using the criteria described under "Comparing a Series of Nested Models" in Miller (2005:342-345). Rewrite that description to rectify any problems you identified.

These types of exercises typically require more time to complete than do the problem sets, so they are best used either as in-class exercises during lab or writing time of at least an hour, or assigned as homework. Additional suggested course extensions can be found in (author citation). The above exercises might also provide ideas for faculty to develop their own suggested course extensions to meet the needs of their students and courses.

SUMMARY

This paper has provided guidelines about where and how to incorporate teaching how to write about multivariate analysis into a graduate curriculum in sociology. Ideally, faculty members will work

together to ensure that these concepts and skills are introduced and reinforced throughout the curriculum, to provide developing sociologists with the opportunity to apply and practice these essential aspects of sociological research.

REFERENCES

Fauth, Rebecca C., Tama Leventhal, and Jeanne Brooks-Gunn. 2004. "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: 2271–84.

Fox, James A., and Marianne W. Zawitz, 2004. *Homicide Trends in the United States*. Washington D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Miller, Jane E. 2005. *The Chicago Guide to Writing about Multivariate* Analysis. The Chicago Guides to Writing, Editing and Publishing. University of Chicago Press.

Miller, Jane E., Paula England, Donald J. Treiman, and Lawrence Wu. "Learning How to Write about Multivariate Analysis for Sociologists: Integrating Concepts from Statistics, Research Methods, and Writing Courses." Working paper, 2009.

Treiman, Donald J. 2009. *Quantitative Data Analysis: Doing Social Research to Test Ideas.* San Francisco: Jossey-Bass.

Zimmerman, David A. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics.* 85(1): 9–23.

Table 1. Regression of cumulative grade point average by own SAT scores and roommate's SAT scores, Williams College classes of 1999–2001

	Coeff.	Standard Error
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 Zimmerman (2003), Table 3.
*: p<.05

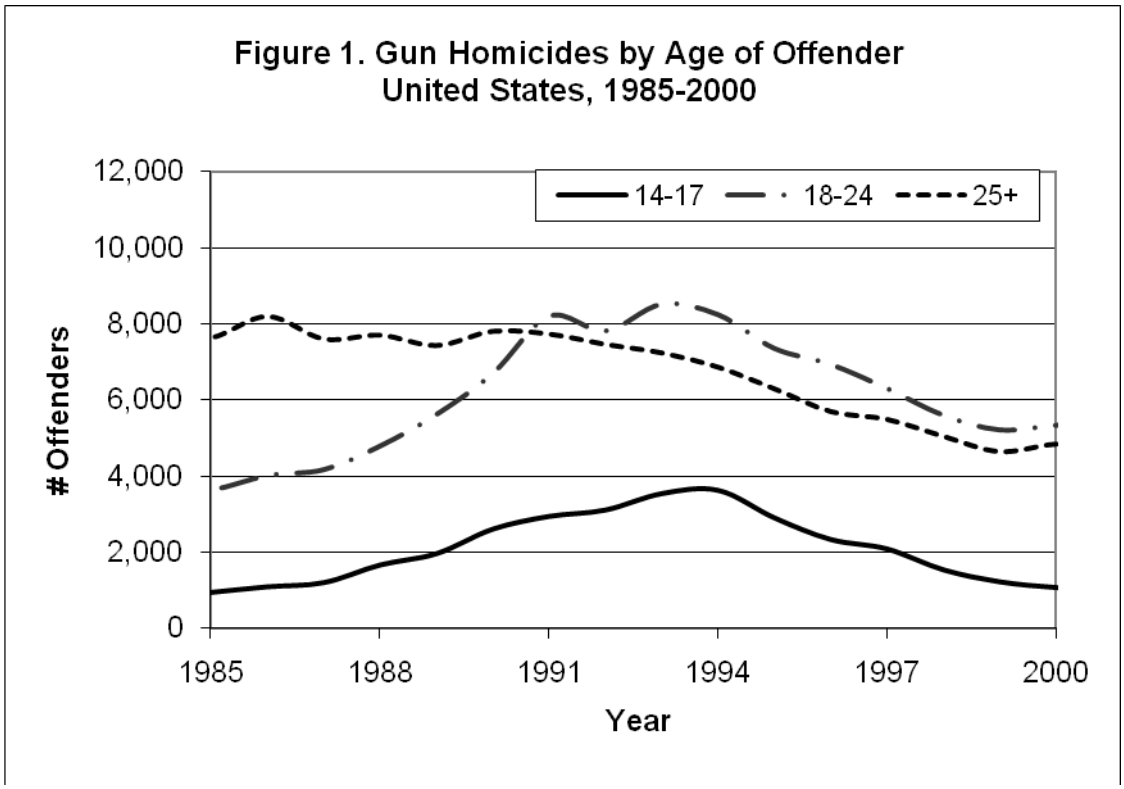
Table 2. Individual background characteristics, neighborhood, and housing characteristics of movers and stayers, Yonkers Residential Mobility Program, 1994–1995

	Residential status			
	Movers (n = 173)	Stayers (n = 142)	Total (n = 315)	χ^2 or F[a]
Background characteristics				
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*
Neighborhood/housing				
Danger (3 items)	0.26	1.29	0.72	144.11***
# of victimizations in past year (1 item)	0.12	0.32	0.21	9.21*
Disorder (5 items)	0.15	1.41	0.72	796.17***
Cohesion (4 items)	0.62	0.40	0.52	43.48***
Resources (5 items)	3.05	2.89	2.98	4.90*
Housing quality (5 items)	0.20	0.54	0.35	54.40***

Source: Adapted from Fauth et al. (2004), Table 1.

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

[a] $\chi 2$ statistic reported for difference in categorical variable between movers and stayers; F-statistic for difference in continuous variable.



Source: Fox and Zawitz, 2004.

APPENDIX: SOLUTIONS TO PROBLEM SETS

Solutions to Problem Set A:

1. In reference to the results shown in Table 1 for the study by Zimmerman (2003):

   a. The dependent variable is cumulative GPA, a continuous variable measured in points, with a theoretically possible range from 0.0 to 4.0.

   b. An OLS model is suitable because the dependent variable is continuous.

   c. The continuous independent variables are student's own and roommate's verbal and math SAT scores, each divided by 100 (see row labels) in the model specification shown in Table 1. Because SAT scores can range from 200 to 800 points, this transformation (change of scale) means that each of these variables could range from 2.0 to 8.0.

   d. The categorical independent variables in the model are gender (ref. = male) and race (ref. = white American citizens, with five dummy variables, one for each of the other racial/citizenship groups [black, Hispanic, Native American, Not a U.S. citizen, Asian]).

2. Sentences to report selected coefficients from Table 1.

   a. "In a study of Williams College students from the late 1990s, each additional 100 points on a student's verbal SAT score was associated with an increase of 0.195 points in cumulative college GPA ($p<.05$), taking into account the effects of race, gender, and roommates SAT scores." *Comment: This sentence incorporates context (who, when, and where), concepts for the independent and dependent variables (verbal SAT and cumulative college GPA), units for each of those continuous variables, direction ("increase"), magnitude (0.195 GPA points), and statistical significance of the association. It also names the other variables controlled in the multivariate model.*

   b. "Female students' GPAs averaged 0.128 points higher than their male counterparts' ($p<.05$)." *Comment: Having already reported the context and conveyed that the results being reported in this paragraph are based on a multivariate model, this sentence does not repeat that information. Again direction ("higher"), magnitude, and statistical significance are reported. Because the independent variable is categorical, the categories are named as they are compared so that the reader can understand the direction of the association; consider the alternative of simply reporting a "gender difference" of 0.128 points.*

   c. "Black students' GPAs were on average 0.264 points lower than those of white students ($p<.05$").

   d. "Compared to Hispanic students, blacks averaged 0.104 points lower on their GPAs ($p<.05$)." *Comment: This sentence requires some behind-the-scenes calculations because the reported coefficients for black and Hispanic students compare each of them against whites. The answer to this "word problem" is reached by subtracting the coefficient for Hispanic from that for black – providing reinforcement of the underlying concept of a reference category and what it means for interpreting a coefficient. Calculating the statistical significance of that difference is another step here; see Miller (2005, pp. 249-250) for more on calculating the standard error of a difference.*

Solutions to Problem Set B:

1. "Chosen by lottery" conveys that respondents were randomly assigned to be either a mover or a stayer. In other words, the study design is a true experiment. It is intended to make the treatment (mover) and control (stayer) groups as similar as possible in terms of all other characteristics, in order to assess the effect of moving *per se*.
2. Residential status (mover versus stayer) is the key independent variable in this study because the main research question to be addressed is whether moving to a low-poverty neighborhood improves outcomes.

3. The neighborhood/housing variables are the dependent variables in this study. They represent six different ways of assessing whether those who move have better outcome than those who remain in their original high-poverty neighborhoods.
4. Background characteristics (including age, gender, ethnicity, educational attainment, household headship, and number of children) are factors that could potentially confound the observed association between residential status and the neighborhood outcomes.
5. In Table 2
    a. Movers were on average just over 2.5 years older than stayers (36.7 years and 34.1 years, respectively, $p<.01$).
    b. The percentage female was virtually identical among movers and stayers.
    c. The average number of victimizations was about three times as high among those who remained in their original neighborhoods as among those who moved to lower-poverty neighborhoods ($p<.05$). Stayers experienced on average 0.32 victimizations, as against 0.12 among movers.
6.
    a. No, the random assignment didn't succeed in equalizing the background characteristics of movers and stayers. "Despite random assignment of treatment and control groups in the Yonkers Residential Mobility Program, there were statistically significant differences in four of the six measured background characteristics between participants who moved versus those who stayed in their original neighborhoods (Table 2). Movers were on average slightly older, more likely to have at least a high school education, less likely to be in female-headed households, and had slightly fewer children than stayers (all $p < 0.05$). No differences were observed in terms of race/ethnicity or gender."
    b. Yes, neighborhood and housing characteristics differed according to residential status. "On all six dimensions studied, outcomes were statistically significantly better among movers than stayers (Table 2). Negative outcomes (danger, victimizations, disorder, and indicators of poor housing) were all lower among movers than stayers, while favorable outcomes (cohesion and resources) were higher among movers than stayers."
    c. These bivariate statistics suggest that a multivariate regression is necessary to assess the impact of residential status on the outcomes studied, net of the potentially confounding effect of the background characteristics. All of the observed differences in background characteristics would be expected to favor better outcomes among movers than stayers regardless of where they live. For example, older age, two-parent households, better education, and smaller families are often associated with better resources than younger, female-headed, less-educated, and larger families. Hence a multivariate model is needed to control for those characteristics in order to measure the net effect of moving versus staying.
    d. Use of observational data on who moved and who stayed in their original neighborhoods would introduce issues of self-selection. People who make the effort to move typically differ in terms of background characteristics factors associated with the outcomes in ways that confound the association between residential status and those outcomes. For instance, people who move are often better educated, higher income, and more motivated than those who stay. An experimental design with random assignment of movers and stayers from among those who apply to move eliminates those sources of confounding, and is thus preferred for evaluating whether moving *per se* improves outcomes.

<u>Comment</u>: Note that these paragraphs each generalize the results rather than writing piecemeal sentences about each of the attributes being compared.