

19. *Speaking about Multivariate Analyses*

SOLUTIONS

1. Figures 19A–19G are slides for a presentation about the physical impact of the planes on the Twin Towers (box 20.3 and figure 20.5).

Annotated example of good writing

- Article from front section of *New York Times*:
 - “First Tower to Fall Was Hit at Higher Speed, Study Finds”
 - E. Lipton and J. Glanz (2/23/02).
- Tailoring to the audience and objectives:
 - An educated lay audience.
 - Two-page article.

Figure 19A.

Airplane speed

- “The FBI said the government's analysis put the speeds at 586 mph for the United flight and 494 mph for the American one.”
 - *Basic principle: report numbers.*
- “In both cases, the planes were flying much faster than they should have been at that altitude the aviation agency's limit below 10,000 feet is 287 mph.”
 - *Basic principle: compare against a standard to help interpret number.*

Figure 19B.

Energy and impact of planes

- “The energy of motion carried by any object, called the kinetic energy, varies as the square of its velocity, so even modest differences in speed can translate into large variations in what the building had to absorb.”
 - *Basic principle: define concepts using simple wording.*
- “That means that while the United jet was traveling only about a quarter faster than the American jet, it would have released about 50 percent more energy on impact.”
 - *Tool: ratio and percentage difference calculations.*

Figure 19C.

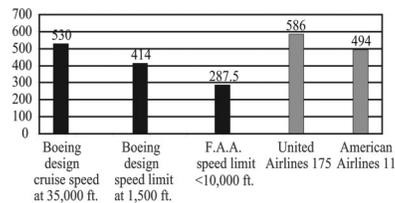
Just how much energy is that?

- “Even at a speed of only about 500 mph, a partly loaded Boeing 767 weighing 132 tons would have created about three billion joules of energy at impact, the equivalent of three-quarters of a ton of TNT.”
 - *Basic principle: interpret numbers and relate them to familiar quantities.*

Figure 19D.

How did speeds compare to design limits?

Impact speed of 9/11 flights* and comparison speeds



* National Transportation Safety Board estimates

- Uses a bar chart to illustrate speed of planes relative to important benchmarks.
 - *Basic principle: choose the right tools.*
- Describe patterns in chart by pointing out that planes' speeds exceeded design limits.
 - *Basic principle: compare against meaningful cutoffs.*

Figure 19E.

Why do design limits matter?

- Such speeds threatened the structural integrity of the planes even before they struck the buildings, because the lower the plane goes, the thicker the air becomes, so the slower the plane must travel to avoid excessive stress.”
 - *Basic principle: explain complex concepts in simple terms, in this case, principles of physics.*

Figure 19F.

Authors' use of tools and principles

- Explained complex ideas without (much) jargon.
 - Energy on impact.
 - Effect of altitude on stress.
- Compared against
 - Useful benchmarks
 - FAA speed limit.
 - Design speed limit.
 - Familiar examples
 - TNT.
- Used appropriate tools.
 - Chart to show relative speed.
 - Prose to:
 - Report a few numbers.
 - Explain patterns.
 - Define terms.
 - Types of quantitative comparisons:
 - Difference.
 - Ratio.
 - Percentage difference.

Figure 19G.

3. Figures 19H and 19I are slides about data and methods regarding CES-D scale for a scientific audience.

CESD scale

- Center for Epidemiological Studies Depression (CESD) Scale
 - Developed by National Institute of Mental Health (NIMH)
- 20 items on frequency of symptoms in past week
 - Each scaled from 0 (“rarely or none of the time”) to 3 (“almost or all of the time”).
- Very good internal consistency:
 - $\alpha = 0.85$ for the general population.
 - $\alpha = 0.90$ for a psychiatric population

Source: Radloff 1977.

Figure 19H.

Factors within the CESD scale

- Four separate factors:
 - Depressive affect.
 - Somatic symptoms.
 - Positive affect.
 - Interpersonal relations.

Figure 19I.

5. Vanna White notes in “[]”, and GEE approach to describe figures or tables
- a. “Figure 6.8 illustrates how the chances of disenrolling from the State Children’s Health Insurance Program vary by reason and

demographic factors, based on a set of competing risks models controlling for all variables shown in the chart. Demographic factors are arrayed across the x axis [wave horizontally]. Each cluster [point to one] shows how that factor is associated with each of the three possible reasons for disenrollment, with other insurance shown in gray, other government program in white, and nonpayment in black [point at respective bars]. The log-odds of disenrollment are shown on the y axis [wave vertically]. Bars that drop below the line at $y = 0.0$ represent lower odds than in the reference category, while those above the line represent higher odds.” (Interpret the pattern as in the description of figure 6.8 on p. 124 of *Writing about Multivariate Analysis, 2nd Edition*.)

- b. “The distribution of federal outlays by major function in the United States in 2000 is shown in figure 6.2b. Human resources (the black wedge [point]) comprised by far the largest single category of federal outlays (61% of the \$1.8 trillion spent that year). The second largest category—national defense (dotted fill)—accounted for only about one-quarter as much as human resources (16% of the total). Net interest, physical resources, and other functions together amounted for the remaining 23% of total outlays [point to each wedge as you mention its category].”
- c. “The predicted pattern of birth weight by race/ethnicity and income-to-poverty ratio (or ‘IPR’) is shown in figure 6.12. The results are based on a multivariate model with controls for gender, mother’s age, educational attainment, and smoking status. The x axis shows the IPR, ranging from 0 to 4 times the poverty line [wave across horizontal axis]. There are separate lines for each of the racial/ethnic groups—the solid line for non-Hispanic whites, the dotted line with triangles for Mexican Americans, and the dashed line with squares for non-Hispanic blacks [point at each in turn, top to bottom on the left-hand side of the x axis]. Predicted birth weight in grams is shown on the y axis [wave vertically].”
- d. “Table 7.1 shows poverty rates for the United States in 1992 under different poverty definitions, for the overall population and several age and racial groups in the rows [gesture at the row labels]. The leftmost column of numbers [point to ‘Current’ column label] is the poverty rate under the current poverty definition, while the next two columns to the right [point to ‘Proposed measure’ column labels] show poverty rates under two alternative definitions. The rightmost two columns [point to ‘Percentage point change’ column label] show the percentage point change in the poverty rate between each of the two alternative definitions and the current definition.” [Note: Explain the alternative poverty definitions on a previous slide, as viewers will focus on the results when presented with the table. Remove the footnote from the slide of this table and turn it into a text slide to precede the table slide.]

“Under either of the proposed alternative definitions, the poverty rate is several percentage points higher than under the current definition. For example, the overall poverty rate would increase by 3.6 points under alternative definition 1, and by 4.5 points under

alternative definition 2 [point to pertinent cells in ‘Total population’ row]. Differences for some subgroups are quite small. For example, the poverty rate for the elderly would be projected to increase by only 1.7 percentage points under alternative 1. For other groups, such as Hispanics, the projected increases are considerably larger: 10.6 points [point to pertinent cell].”

- e. “Figure 16.1 shows the predicted differences in birth weight by mother’s educational attainment and race/ethnicity. Racial/ethnic groups are shown in the legend [name them and point to associated bar colors: white bars for non-Hispanic white infants, striped for Mexican American, and black bars for non-Hispanic blacks]. Educational attainment is shown in increasing order across the *x*-axis [name them and gesture along the horizontal axis]. The length of each bar shows the difference in predicted birth weight (grams) between the pertinent group and non-Hispanic whites with at least some college, which is the reference category from the multivariate model.” (Describe the pattern as in box 16.1 on p. 363 of *Writing about Multivariate Analysis, 2nd Edition*.)

7. Slides to present results of Yonkers Residential Mobility Program evaluation (Fauth et al. 2004).

- a. Figures 19J through 19L are slides demonstrating why a multivariate model is needed.

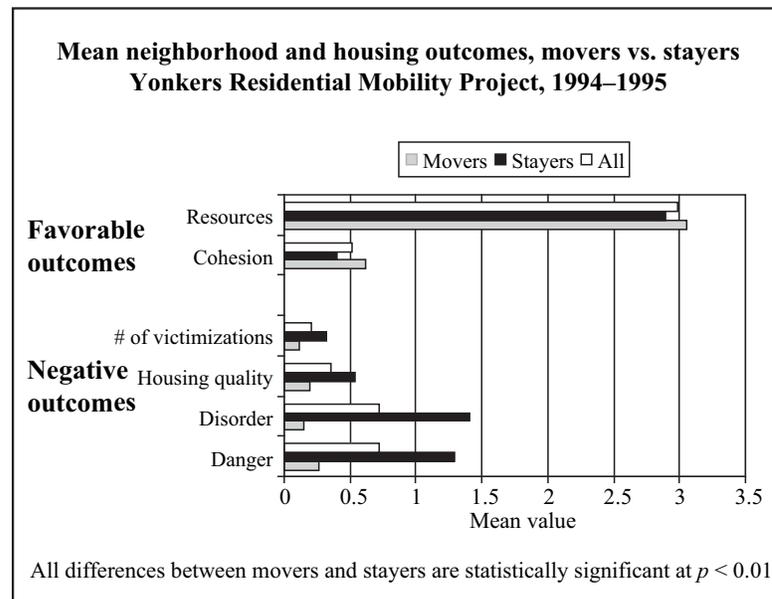


Figure 19J.

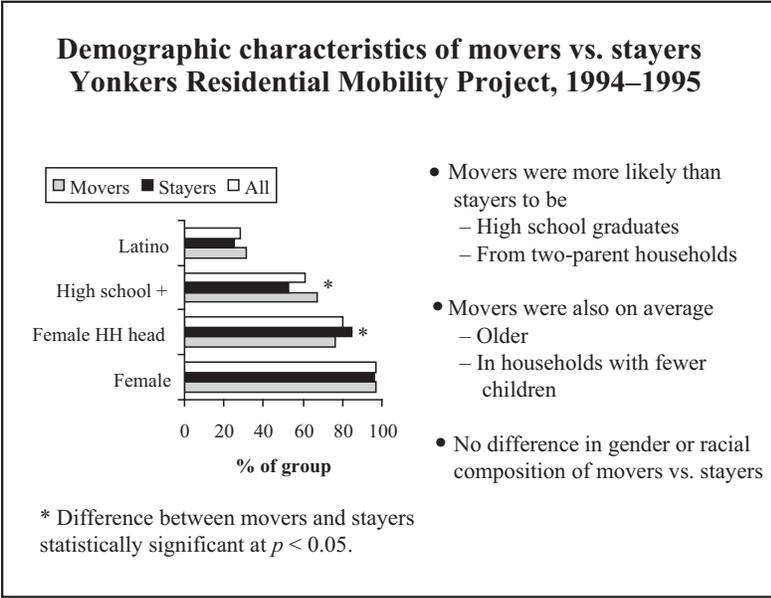


Figure 19K.

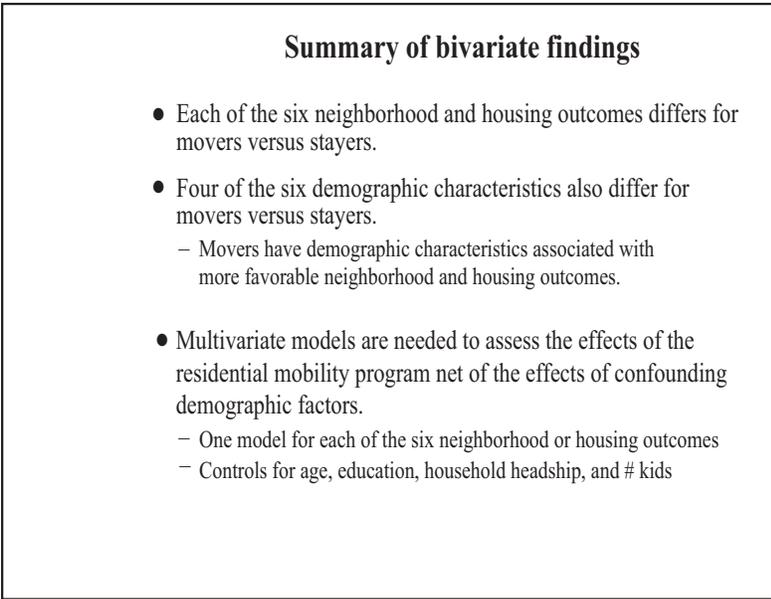


Figure 19L.

b. Figures 19M and 19N are slides presenting multivariate model results.

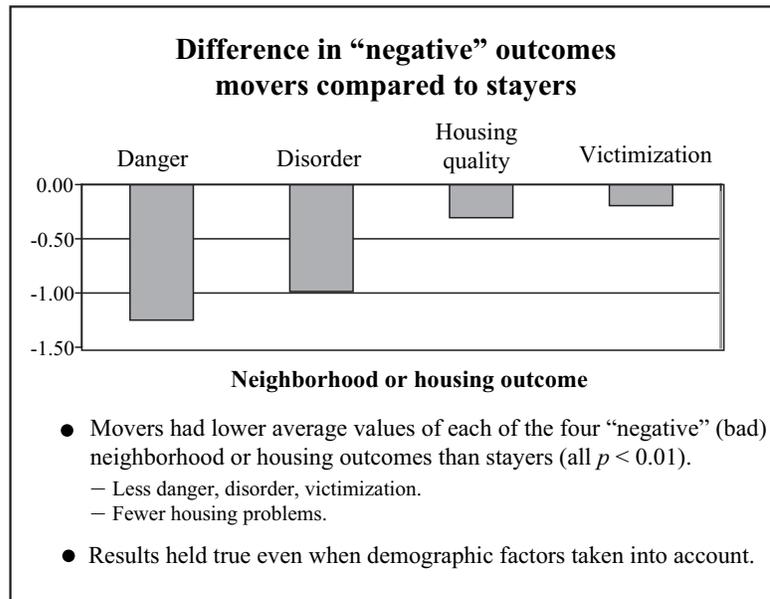


Figure 19M.

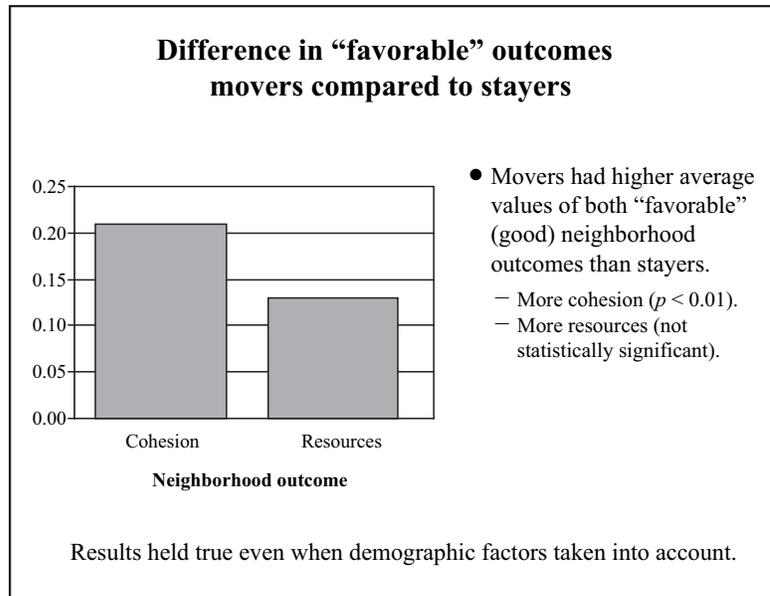


Figure 19N.

c. Vanna White notes are shown in “[]”

“Figure 19J shows mean values of six different measures of neighborhood and housing quality for low-income families who moved to low-poverty neighborhoods versus those who stayed in high-poverty neighborhoods. In all six dimensions studied, outcomes were statistically significantly better among movers (the gray bars) than among stayers (the black bars). Favorable outcomes (cohesion and resources) [gesture at top two clusters on the chart] were higher among movers than stayers, whereas negative outcomes (danger, victimizations, disorder, and indicators of poor housing) [gesture at four bottom clusters on the chart] were all lower among movers than stayers.

“However, it is important to consider whether differences in demographic characteristics might explain some of the observed differences in these outcomes. Although participants in the Yonkers Residential Mobility Program were randomly assigned to be movers or stayers, some differences in these characteristics are possible. In figure 19K, we see that four of the six background characteristics are more auspicious among movers than stayers. Movers were more likely to be from two-parent households and to have completed high school [point to respective clusters on chart]. They were also on average older and had fewer children in the household.

“[Transition to slide 19L] These bivariate statistics suggest that multivariate models are needed to assess the impact of residential status on each of the outcomes, 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 residence. 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 multivariate models are needed to control for those characteristics.

“Figure 19M shows results of multivariate models of the four negative measures of neighborhood characteristics and housing quality studied as part of the Yonkers Residential Mobility Program (danger, victimization, disorder, and housing problems) [point to respective bars, moving left to right across x -axis]. Even when the effects of potential confounders were taken into account, subjects who moved had statistically significant better values of each of these four outcomes than those who remained in their original neighborhoods. Put differently, movers experienced less danger, victimization, disorder, and housing problems than stayers.

“Figure 19N shows the results of multivariate models of the two favorable outcomes (cohesion and resources). Both were higher (better) among movers [gesture along y axis], but the difference in resources was not statistically significant. Although some of the background control variables were statistically significantly associated with one or two of the outcomes, none showed a consistent pattern of association.”