5

Creating Effective Tables

1. Title for table 5A: "Median age of the U.S. population, 1960 to 2000."

3. Notes to table 5B.
   Spell out BMI (body mass index), show the formula, and provide a citation.
   Specify numeric cutoffs for income or the income-to-poverty ratio to define "poor," "near poor," and "nonpoor."
   Define what * and ** denote.
   Cite the data sources.

5. Design tables for the given topics.
   a. Title: "Age, gender, race, and educational attainment composition of [fill in who, when, and where for study sample]." Table structure: Demographic variables in the rows, with units specified in row header for age, subgroups for the categorical variables shown with indented row headings. Columns for number of cases and percentage of cases. Note citing data source.
   b. Title: "Pearson correlation coefficients between height, weight, percentage body fat, systolic blood pressure, and resting pulse, [W's]." Table structure: one row and one column for each variable, with label indicating units or footnote callout for abbreviated units. Correlations reported in the below-diagonal cells (see *Writing about Multivariate Analysis,* table 5.7, p. 98, for an example). Symbols in the table cells to identify $p < 0.05$, with a note to explain the meaning of the symbol. Another note to define unit abbreviations
   c. Title: "Estimated odds ratios and z-statistics from a logistic regression of high school graduation, by gender and residence, United States, 1998." Mother's and father's educational attainment and occupation in the top rows, followed by other independent variables. Column spanner for each gender over columns for urban and rural (total of four models), with z-statistics in parentheses below odds ratios for each independent variable with symbols denoting $p < 0.01$ and $p < 0.05$. Goodness of fit statistics and degrees of freedom for each model in rows at bottom of the table. Footnotes citing data sources, symbols.
   d. Title: "Low, medium, and high projections of number of college degrees earned (thousands), by region, United States, 2010 to 2025." Columns for low, medium, and high with a spanner labeled "scenario," rows for years. Notes about data sources, assumptions used in each scenario.

e. Title: "Net effects of an interaction between student's high school class rank and mother's educational attainment on student's first-year college grade point average, high school classes of 1995 to 2000." One column each for bottom, middle, and top tercile of class rank with a column spanner labeled "class rank," one row for each level of mother's education (<HS, =HS, >HS). Interior cells include estimated values of first-year college GPA to nearest two decimal places with symbols denoting statistical significance. Notes specifying data source and other variables controlled in the model (or naming a table in which those estimates are shown), identifying the top terciles as the reference category, and defining symbols used to denote statistical significance.

7. Errors are labeled in the table using lettered superscripts keyed to the comments below.

Table 5F. Results of a logistic regression of political party preference,[a] Whatnation, 2004

Variable	Odds ratio	Confidence interval[b, c, d]	Wald chi-square
Age group 2[e]	1.82	−0.015–3.83	4.13
Age group 3	2.01	−0.25–5.19	3.67
Race[f]	0.53	−1.31–1.03	5.99
Proportion poor[g]			
<10	1.26	−0.51–2.64	0.67
10–19	2.36	0.04–5.36	7.25
20–29[h]			
>29	0.35	−2.02–0.93	7.69

Comments on errors in table 5F:
a. The category of the dependent variable being modeled is not specified, so it is unclear whether the regression is estimating relative odds of a Democratic party preference or a Republican party preference.
b. The width of the confidence interval isn't specified. (The correct value is 99% CI.)
c. The confidence intervals are specified in terms of log-odds, not odds ratios. (You can tell because odds ratios can never be below 0, but the corresponding log-odds will be <0.0 whenever OR<1.0.) Either report log-odds instead of odds ratios and keep the current CI, or calculate the CI in terms of odds ratios.
d. Using a dash ("–") to separate confidence limits that include negative values is confusing. Replace the dash with a comma, e.g., −0.015, 3.83
e. The reference category for age group isn't included in the table, and the labels for the other age groups don't provide enough information for readers to infer the identity of the reference category.

f. The identities of the included and reference categories of the race dummy variable cannot be determined by the row label "Race."

g. Proportions must be between 0.0 and 1.0, therefore the reported values are probably percentages. Either change the label to read "Percentage poor," or convert the values to proportions and label accordingly (e.g., < 0.10, 0.10−0.19).

h. The reference category could be more clearly marked using one of the conventions mentioned on page 87 of *Writing about Multivariate Analysis*. Identify the convention with a note to the table.