Chapter 11
Presenting Results of Statistical Analysis

This chapter is intended to be a practical guide to help with the construction of tables and figures and with the general presentation of results of statistical analysis in a research paper. Constructing tables and figures well and writing a results section so that it appears to make a coherent point—and does not wander—is as important as constructing a solid research question and conducting analyses correctly to answer it. That is, if the reporting of the results does a poor job of telling a story that can answer the research question, it is ultimately pointless to have developed a good literature review and research question. For that matter, the statistical analyses may have been performed extremely well, but if the results are not displayed in a way that is easy to see and understand, the analyses have been a waste of time.

The data analysis process should, if it is done well (i.e., the data have been thoroughly investigated), yield far more results than one can possibly report in a single paper. Part of the art of writing a good paper is determining what should be reported (and in what order) to make a good story. I recommend that, before one begins to actually write the results section, one should (1) decide what results will be reported, (2) make tables and figures and put them in the order to be discussed, and (3) plan to write the results from the tables and figures.

11.1 Making Tables

Tables are the key way that results of statistical analyses are presented. A table may present descriptive statistical information, or it may report results of inferential analyses and statistical tests. In general, tables should be easy to read, should not contain unnecessary information, but should contain all information relevant to understanding the analyses, and they should be able to stand alone from the text. That is, one should not have to reference the text of a paper to understand a table. With these ideas in mind, there are a number of important rules for making tables. Some of them are guidelines more than rules; different journals, researchers, and disciplines have different styles.

S.M. Lynch, *Using Statistics in Social Research: A Concise Approach*,
DOI 10.1007/978-1-4614-8573-5_11,

	Mean or Percent (s.d.) [range]	
Variable	**Male (n = 888)**	**Female (n = 1148)**
Age	47.8(17.0)[18,89]	48.1(18.1)[18,89]
Nonwhite	22.6 %	25.3 %
Education (yrs.)	13.4(3.3)[0,20]	13.5(3.1)[0,20]
Income ($1000)	31.0(31.1)[0,119.6]	24.4(27.5)[0,119.6]

Note: Income statistics include persons who have income of 0.

Table 11.1. Descriptive statistics for variables used in analyses by sex (2010 GSS).

First and foremost, *do not* simply cut and paste raw output from a statistics software package. Raw output from a statistics package does not constitute a table and is insufficient as the body of the table. The material is not easy to read, and there is usually more material reported than is necessary for discussion. Furthermore, the material is usually not presented in the most readable format.

Second, tables must have titles, and the title must fully describe what the table contains. It is insufficient to have a title like: "Table 11.1. Results." Instead, the title should be detailed: "Table 11.1. Results of Regression Analyses of Income on Education and Background Variables (NHANES sample, n = 3,076)." If there is too much information to put in the title, add a footnote. For example, in the previous title, I may wish to footnote the table in order to define "NHANES."

Third, beyond any additional information that cannot be fit into the title, most tables of results have footnotes reporting symbols for levels of statistical significance. The usual convention in social science research is: "# $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$."

Fourth, tables generally consist of columns. The first column is almost always a list of the variables for which descriptive statistics are being reported or for which regression coefficients are being reported. Subsequent columns are used for (1) the regression coefficients or descriptive statistics, and (2) results from additional models (if more than one model is being reported in the table) or results for different groups that were modeled (e.g., races, sexes, etc.).

Fifth, variable names should be recognizable. Do not use variable names like "v2103," or whatever variable name existed in the data set; instead, use names like "Male" or "Nonwhite." In fact, for dummy variables, name them after what they indicate, as with "male;" do not use "sex," because it is then necessary to define in a footnote what value is 1 vs. 0.

Sixth, usually, the regression coefficients (or means or percents, if the table is reporting descriptive statistics), are reported first, followed by the standard errors of the coefficients in parentheses (or standard deviations, if the table is reporting descriptive statistics).

Finally, given that virtually all tables have footnotes, I recommend the following format for any table: (1) Title in bold, (2) one or two horizontal lines, (3) main

	Male (n = 807)		Female (n = 994)	
Variable	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Intercept	−20.95***	−20.43***	−27.96***	−34.94***
	(5.22)	(5.72)	(4.58)	(5.14)
Age	.09	.09	.08#	.08#
	(.06)	(.06)	(.05)	(.05)
Nonwhite	−5.82*	−7.96	−8.83***	13.82#
	(2.42)	(9.79)	(1.83)	(7.91)
Education	3.85***	3.81***	4.00***	4.51***
	(.31)	(.35)	(.27)	(.32)
Nonw*Educ		.17		−1.70**
		(.74)		(.58)
R^2	.18	.18	.22	.22

# $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

[a] The sample was reduced to individuals who had non-zero earnings.

Table 11.2. Results of multiple regression analyses of income on race and education by sex (2010 GSS)[a].

body of the table, (4) one horizontal line, (5) the p-value symbol note, (6) additional footnotes. The title of a table may, alternatively, be placed at the bottom of the table as done throughout this book.

Tables 11.1 and 11.2 illustrate table construction. Table 11.1 reports descriptive statistics; Table 11.2 reports the results of a set of regression models. These tables demonstrate the rules/guidelines shown above. The title of each table describes what the table contains, including reference to the sample. Here, the sample name (GSS) is an abbreviation and should probably be spelled-out in a footnote, but in most journals that would publish a paper using these data, the abbreviation would be known by readers. Both tables contain a considerable amount of information. Table 11.1 provides detailed summary statistics by two sex groups in the sample for four variables.

Table 11.2 provides regression coefficients for multiple regression models predicting income conducted for each sex. In one model for each sex (Model 2), an interaction between race and education is included; in the other model (Model 1) only main effects are included.

11.2 Making Figures

In addition to tables, figures are an important means for presenting interrelationships between concepts in a theory (a "conceptual plot") and presenting results of analyses or showing patterns, trends, and relationships between variables in data. The most common data-based figures used in social science papers include histograms, bar charts, scatterplots, and line plots. Histograms display the frequency distribution of a variable (see earlier chapters). Bar charts represent means or totals of some variable for different groups. The x-axis of these plots references the different groups, while the y-axis (the height of the bars) is the mean or total of the variable of interest. Scatterplots are used to plot one continuous variable against another so that any relationship between them can be seen. Finally, line plots are often used to show trends in variables across time. They are also often used after a regression analysis to show model-predicted patterns/relationships between variables. Each of these types of figures, along with some others, has been used throughout this book.

A good figure is one that conveys considerable information very simply and does not mislead a viewer. To that end, there are a number of rules for making figures. Some of the rules are the same as for tables, including that a figure must stand alone without relying on reference to the text. There are a number of additional rules for constructing good figures.

First, as with tables, plots that are produced in statistics packages using default settings do not constitute legitimate figures. They will usually violate other rules that we will discuss. Instead, the data for figures will generally need to be entered into Excel or some other package that is specifically geared for producing figures.

Second, figures must have titles, and the title must fully describe what the figure contains. If plotting one variable against another, the general rule for phrasing is that Y—the variable on the y-axis—is plotted against X.

Third, variable names should be recognizable in the figure, just like they should be in a table. Don't use variable names like "v2103," or whatever variable name existed in the data set; instead, use names like "Male" or "Nonwhite."

Fourth, label all axes with appropriate labels. Do not leave this to the default settings of the software package.

Fifth, choose the scale of the axes appropriately. Do not truncate the domain or range of the data. The origin (0,0) should almost always be included to prevent distortion of the relationship between the variables in the plot. Do not leave the choice of domain or range of axes to the default settings of the software package.

Finally, do not make the figure more complex or fancier than it needs to be to display the material. For example, bar graphs do not necessarily need to have colored bars, nor do the bars need to be three-dimensional. In scatterplots, plotting characters do not generally need to be more complex symbols than dots or circles. In brief, a good figure is one that is not confusing, not one that has a pretty or sophisticated appearance.

Figures 11.1 through 11.3 highlight some of these rules. All three figures use the same data, but they appear very different and have different levels of readability.

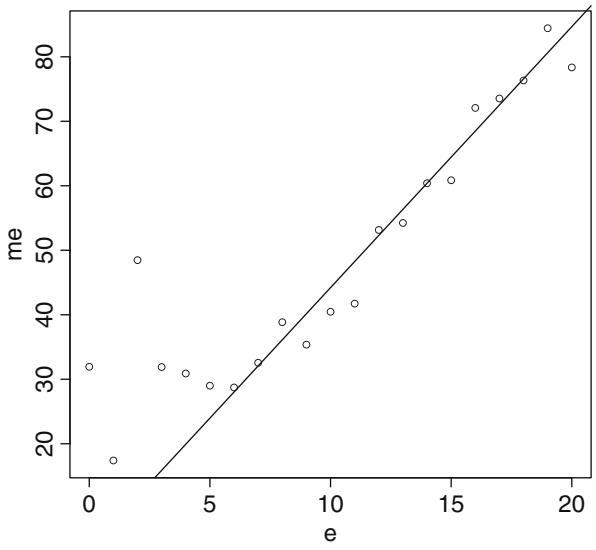

Fig. 11.1. Income and schooling

Figure 11.1 shows the relationship between education and income by plotting mean income for persons at each level of education—measured in years—in the GSS. The plot is the default plot obtained using my preferred software (R). This figure violates almost every rule for making figures. The title of the figure is insufficient to understand what the figure shows. What are the circles in the figure, and what is the line? Where did the data come from? The axes are labeled poorly: what are "e" and "me"? Finally, the image itself provides a somewhat misleading impression of the strength of the education-income relationship. Notice that the y-axis in the plot starts at 20, rather than 0. Thus, the y dimension of the plot is compressed, making the slope of the line in the figure appear quite steep.

Figure 11.2 remedies almost all of the problems that were present in Fig. 11.1. The title is much more detailed. The axis labels have been replaced with words, so that it is very clear that the x-axis refers to years of schooling, ranging from 0 to 20, while the y-axis refers to mean income. The label clarifies that the units are thousands of dollars. Importantly, the axis range now starts at 0, and the axis has been expanded upward to 100 (again, in thousands). Obviously, we could increase the upper limit further, but all of the data points seem to fall pretty evenly between 0 and 100, and 100 is a reasonable upper limit for income in the US population. Very few earn more than this. Finally, the inclusion of a legend clarifies that the observed means in the data are represented by the circles, while the line is a predicted regression line, showing the smoothed expected relationship between years of schooling and earnings. Although not shown in the figure, each year of schooling appears to be associated with a $4,000 increase in income.

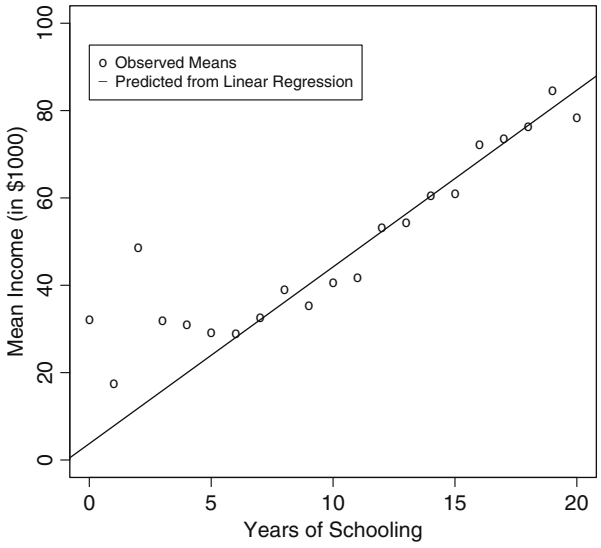

Fig. 11.2. Observed and predicted mean income by years of education (n = 12,924, 1972–2006 GSS data for persons ages 30–54; income in 2006 dollars).

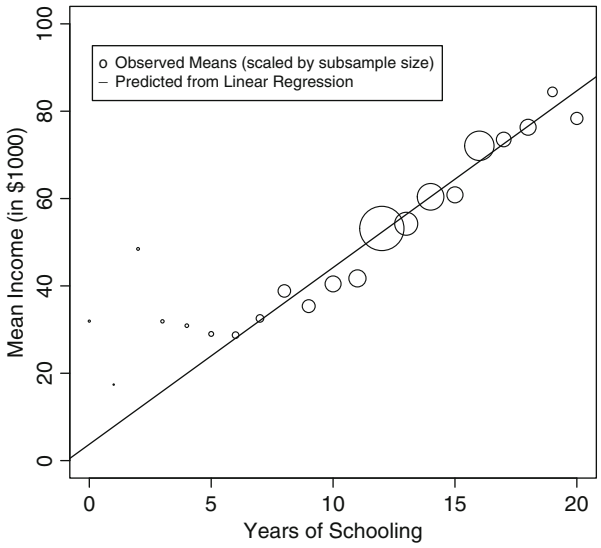

Fig. 11.3. Observed and predicted mean income by years of education (n = 12,924, 1972–2006 GSS data for persons ages 30–54; income in 2006 dollars).

One problem that remains in the figure is that the regression line seems to fit the observed means somewhat poorly, especially at the lowest levels of education. Figure 11.3 clarifies why: There are very few individuals in the sample at these

levels of education, and so the regression results are driven primarily by those with more schooling. The final figure demonstrates this by using circles at each level of schooling that are sized based on the number of individuals in the sample at that level of schooling.

11.3 Writing the Results Section

Once tables and figures have been created and placed in the order in which they are to be discussed, writing the results section of a paper is relatively simple. Remember that tables and figures must stand alone; that is, they must be able to be viewed without reference to the text. Similarly, the results section of the paper must be readable without reference to the tables and figures. This does not mean that one cannot refer to the tables and figures in the text—indeed, they should be referenced. It simply means that a reader must be able to understand the results section without looking at the tables and figures. To some extent, then, the results section and the tables and figures may seem somewhat redundant. However, the tables and figures are provided to supplement the text and give a quick, visual summary of the writing in the results section.

I recommend printing each table and figure on a separate page, placing them in order, and then writing the results section as if you were explaining the tables and figures to someone verbally, walking through the table and figure results one step at a time, following the storyline. Every single numerical result that is presented in the tables and figures need not be tediously addressed. Instead, focus on the main highlights. For example, in presenting the results from Table 11.2 (the regression model results), the intercepts in each model do not need to be discussed, nor do the standard errors. Instead, something like the following is sufficient:

> The regression model results presented in Table 11.2 show that older persons uniformly tend to earn slightly more than younger persons, although the results in each regression model are either not statistically significant or only marginally so. For example, among men, Models 1 and 2 show that each year of age produces an increase of \$90 in earnings, although the result is not statistically significant. Among women, each year of age produces an expected increase of \$80 in earnings, but this result is only marginally significant ($p < .1$). For both sexes, Model 1 results show that each year of schooling is worth approximately \$4000 more in earnings (\$3850 for men; \$4000 for women). Model 1 also shows that nonwhites earn less than whites. The race difference in earnings is greater for women than for men, however. Nonwhite men earn about \$6000 less than white men, while nonwhite women earn about \$9000 less than white women.
>
> The results of Model 2, in which an interaction between race and education is included, show that the returns to education for whites and nonwhites vary across sex. For men, the Model 2 results are very similar to the Model 1 results. The interaction

between race and education is only .17 and is not statistically significant. For women, the Model 2 results are very different than the Model 1 results. In Model 2, nonwhite women with no schooling earn *more* than white women with no schooling, but that advantage erodes with more schooling. Given the drastic change in coefficients from Model 1 to Model 2, some predicted values for earnings provide a clearer picture. For men at age 45 (age is constant for all predictions), whites with 12 years of schooling earn approximately $30,000 on average. The average is roughly double for those with 20 years of schooling. For nonwhite men, average earnings are about $21,000 at 12 years of schooling (a difference of $9000 with white men with the same level of schooling), but about $55,000 at 20 years of schooling—a difference of only $5000 with white men with the same level schooling. Thus, each year of schooling nets slightly more for a black male than a white male.

For women, this pattern is reversed. White women with 12 years of schooling earn approximately $23,000 on average, and that average more than doubles to $59,000 at 20 years of schooling—almost as much as men with 20 years of schooling. For nonwhite women, average income is $37,000 for those with 12 years of schooling, but the average only increases another $2000 for those with 20 years of schooling.

Overall, the results indicate that nonwhite men gain slightly more per year of schooling than white men (albeit not a statistically significant amount), and income increases substantially with education. In contrast, while white education at low levels of education earn far less than white males, nonwhite women earn considerably more than nonwhite males at low levels of schooling. However, white women gain far more from schooling than nonwhite women, approaching the earnings of their male counterparts, while nonwhite women appear to gain little from additional schooling, thus earning substantially less than nonwhite males.

Based on how I wrote this description of the table, it is apparent that my research question is/was focused on whether race differences in the return to education varied by sex (a three-way interaction). If my focus had been simply on education's effect by sex, I would have focused on discussing the fact that the Model 1 results show that men and women appear to receive roughly the same returns to schooling (coefficient of 3.85 vs. 4.00), and I would have excluded Models 2.

Notice how, in this write-up of the results of the table, I did not discuss every number in the table. Results sections of papers are often inherently boring to read, and so you should focus on the key results to keep the reader interested and to keep the focus on the story you are telling. Also notice how I did not make any value judgments about the results, and I did not draw any implications regarding my theory and research question: This discussion is for the "Discussion" section of the paper. The results section simply presents the results.

11.4 Writing the Discussion Section

The Discussion section of the paper summarizes the results further and discusses how the results answer the research question, and more broadly, what the implications of the results are for the larger theory from which the research question was drawn. In this particular example, I may wish to say something like:

> Stratification research has long focused on both racial and gender differences in earnings (citations), as well as on the importance of education to earnings (citations), and to a lesser extent the differential returns to education for men versus women and whites versus persons of other races (citations). However, little research has simultaneously considered that gender returns to education may themselves vary by race. Yet, such a consideration is theoretically important (why?; citations)... In this study, we found that there is a substantial difference by sex in how education affects earnings for whites and nonwhites. While white men at low levels of schooling earn considerably more than nonwhite men at comparable levels, the earnings gap shrinks substantially across levels of schooling. The pattern is much different for women. Nonwhite women earn more than white women at lower levels of schooling, but their wages are relatively flat across education. White women, in contrast see large returns to each additional year of schooling so that they earn nearly as much as white men at the highest levels of schooling. All in all, the results support a "double jeopardy" hypothesis: that being both female and a member of a racial minority present a double threat to earnings that education simply does not overcome. The implications of these findings include that research on discrimination, and stratification research more broadly, should not focus on sex or race alone; instead, research should consider the nexus of these two social groupings as suggested by intersectionality theory...

Notice how, in this excerpt, I have not discussed particular, numerical results like I did in the Results section. Also notice how I placed my findings (and my study more generally) in the context of previous research, and I have drawn implications for subsequent research. (as a side note, research in stratification has already considered these issues; this is simply an example).

The end of the discussion section usually points to limitations of the present research presented in the paper, and it offers suggestions for future study. Sometimes this information is presented in a separate Conclusions section; sometimes the paper ends at the end of the Discussion. In this example—just to provide one of many shortcomings—I might point out that I did not control on employment status other than to exclude persons who reported no income. It is quite possible that differences by sex and race in the proportion engaged in part time work explains some of the findings, and this limitation should be mentioned (if not corrected in the original analyses!)

11.5 Conclusion

Making good tables and figures is crucial to writing a good research paper. In this chapter, we have discussed some basic rules for constructing good tables and figures, as well as writing summaries of them. At this point, we have covered the entire process of conducting basic quantitative research, from developing a research question, to constructing a survey instrument, to obtaining data, to analyzing it using basic statistical methods, to reporting the results. You should therefore be ready to conduct your own quantitative research from start to finish. In the final chapter, I offer some suggestions for additional reading to help flesh out each of the topics we have discussed throughout the book.