PLSC 502: "Statistical Methods for Political Research"

How To Read Tables and Figures
August 30, 2016

Introduction

We'll spend the first day discussing something you'll be doing a lot over the next four (forty?) years: reading and understanding statistical results, as presented in tables and figures. We'll do so by going over two examples, one older and one more recent: Rosh (1987) and Gomez, Hansford, and Krause (2007).

Tables

For better or worse,[1] social scientists turn first to tables to present empirical/quantitative data and findings. Tables come in all shapes and sizes; a few key things about tables, in general, include:

- **Content**. What is the author presenting with the table? Data? Coefficient estimates? A Model? Predictions?
- **Organization**. Columns and rows, of course, but always keep in mind what each one is.
- **Role**. What point does the table make? How does it fit into the larger argument or set of arguments that the author is trying to make?

There are three general types of tables that are very common in social science research;[2] we'll discuss them in decreasing order of prevalence.

Tables of Coefficient Estimates

The most common tables found in social science articles present *estimated quantities of interest* (t-tests, regression coefficients, etc.) of some sort. The "usual" organization for such a table:

- Lists a *title* at the top,

[1] For the "worse" perspective, see Gelman, Andrew, Cristian Pasarica, and Rahul Dodhia, 2002, "Let's Practice What We Preach: Turning Tables Into Graphs," *American Statistician* 56:121-130, and/or Kastellec, Jonathan P., and Eduardo L. Leoni, 2007, "Using Graphs Instead of Tables in Political Science," *Perspectives on Politics* 5(4):755-71.

[2] And of course there are other types as well; but these three constitute (probably) 98 percent of all published tables in social scientific work.

- Has a list of (independent) *variables* in the far left-hand column,
- Has one or more columns which denote separate (empirical) *models* that the author has estimated,
- Has cell entries that present the relevant *estimates*,
- Includes some indication of the *variability* of those coefficient estimates (e.g., standard errors or "*t*-scores"),
- Often includes *asterisks* (*) to indicate levels of statistical significance of those estimates, and
- Includes *notes* and other ancillary information at the bottom of the table.

Each of these components tells us something different about what the author has done; we'll get into more details in the examples below. In general:

- Positive coefficient estimates mean that positive changes (increases) in that X variable are associated with positive changes (increases) in the phenomenon of interest (Y).
- Negative estimates indicate the reverse.
- Coefficient estimates that are much larger than their standard errors are considered more "reliable" ("statistically significant"). This will be apparent by one or more of:
    - Standard errors (usually in parentheses) that are more than twice as small as their associated estimates, and/or
    - "*t*-scores" that are equal or greater than two or so, and/or
    - Asterisks ("stars") indicating statistical significance.

Tables of Predictions

Authors often also present tables of *predictions*; these are typically used to convey the substantive significance of the author's findings. While they vary somewhat, tables of predictions usually include:

- A left-hand column that shows different values for one or more key independent variable(s) in the model,
- One or more columns that include predictions of the dependent variable at the values of the independent variables, and
- Occasionally, some sort of measure of uncertainty of those predictions.

There are no tables of predictions in our examples, but we'll discuss them when we get to the Gomez et al. article anyway.

Tables of Data

Finally, authors sometimes present their actual *data* in tabular form. This practice was more prevalent in olden days (e.g., the 1980s) than it is now, but we still see it done once in a while. We'll discuss data at length next class, and see an example of this practice in the Rosh article (below). For now, bear in mind three general types of such tables:

1. *Tables of actual data*, that is, tables that simply show the units of analysis and values of the variables for each (as in the Rosh article).
2. Tables of *summary statistics*, like means, standard deviations, etc., that describe the data.
3. *Crosstabs*, which show how the data are distributed across one or two (or perhaps three) key variables at a time.

Figures

"Figures" refer generically to graphical presentations of information; here, we're most interested in statistical information. We'll spend more time later in the course learning about different types of figures and how to use them; for now, what we care about is reading and interpreting them.

I won't go into all the different types of figures here. Instead, here are some things to keep in mind as you're reading a figure:

- **Content**. Again: Data? A Model? Predictions? Coefficient estimates? Equilibria (as in game theory)?
- **Organization**. This can take on a bunch of different forms, all of which we'll start discussing a bit later.
- **Role**. As with tables: What point does the figure make? How does it fit into the larger argument or set of arguments that the author is trying to make?

General Tips/Hints for Reading Tables & Figures

1. *The title and notes can be illuminating.*
2. *Location.*
3. *Don't get too bogged down in details.*

Readings: Rosh (1987)

Points:

- Figure 1: Scatterplot, with the two key variables of interest.
    - The Y-axis is the thing he wants to explain,
    - The X axis is the key variable doing the explaining,
    - The data points are arrayed at their respective values on both variables.
- Table 1: Data (the actual data he uses)
- Table 2: Regression results
    - Column of independent variables
    - "Standardized" coefficient estimates
    - Unstandardized estimates
    - Standard errors, and
    - P-values (statistical significance)

Readings: Gomez, Hansford, and Krause (2007)

Points:

- Figure 1 is a comparative map. Maps are great.
- Table 1 is a standard table of regression estimates, for two separate models of turnout.
- Figure 2 is a plot (histogram) of predictions. (Q: How could this have been a table?)
- Table 2 is a second table of regression estimates, this time for the GOP-share variable.
    - Tests two alternatives: a "uniform" effect vs. one where the effect of weather is contingent on the fraction of Republicans in the electorate.
    - Finds support for the latter...

Finally: See the "Glossary" on the **ANGEL** Site