How To Read A Regression Table:
A Quick, Completely Non-Technical Guide

Jane L. Sumner
Department of Political Science
University of Minnesota, Twin Cities

Political science students often, at one time or another throughout their education, encounter a regression table. Buried deep somewhere in the last half of an article assigned for class, the regression tables take up a page or more -- sometimes requiring the reader to turn the page to take in the numbers in all their glory. It looms, large and intimidating, inviting the student to quickly flip past. Yet the regression table can tell you many important things, including some that are not always described in the text, so readers ignore the tables at their peril. In this brief guide, I provide the basic information students need to read and interpret just about any regression table, without needing to understand any of the underlying statistics.[1]

The table below, labeled "Table 3. Sheriff Immigration Policies" from Farris and Holman (2017) is one such regression table. It looks like most other regression tables. It has words at the top and on the left side. It's full of numbers. Some of the numbers have symbols after them, and some of the numbers are in parentheses. In the sections that follow, I'll decompose the table and explain what we can learn from it -- even without reading the article itself.

1. **Dependent Variables**

Across the top row, there are five groups of words: Immigration Attitudes Factor, Number of immigration checks, Number of immigration checks (w controls), Policy on immigration status inquiry, and Policy on immigration status inquiry (w controls). These are the five *dependent variables*. The dependent variable is the thing we are trying to explain -- like, for instance, a sheriff's attitude toward immigration, or the number of immigration checks. The type of statistical model the researchers chose to use often depends upon the values these can take. A dependent variable that can be anything from 0 to 100 requires different statistical tools to understand than one that can only be 0 or 1. But ultimately, you don't need to know which model they chose in order to understand the basics of what the table tells you. You do need to know that each of these columns is something the researchers are trying to explain, and the words at the top should give you some idea of what that thing is. Immigration Attitudes. Number of Immigration Checks. And so on.

[1] There are plenty of good, more technical guides on the internet that will get into the specifics of different models, discuss data and assumptions, or dig deeper into how to put together a regression table. This is not that guide. If that's what you're looking for, I suggest Steven V. Miller's "Reading a Regression Table: A Guide for Students" (http://svmiller.com/blog/2014/08/reading-a-regression-table-a-guide-for-students/), Abby Long of EGAP's "10 Things to Know About Reading A Regression Table" (https://egap.org/methods-guides/10-things-know-about-reading-regression-table), or Sharad Vijalapurm's (freeCodeCamp) "How to read a Regression Table" (https://www.freecodecamp.org/news/https-medium-com-sharadvm-how-to-read-a-regression-table-661d391e9bd7-708e75efc560/).

Table 3. Sheriff Immigration Policies.

	Immigration Attitudes Factor	Number of immigration checks	Number of immigration checks (w controls)	Policy on immigration status inquiry	Policy on immigration status inquiry (w controls)
Sheriff characteristics					
Immigration Attitudes Factor	—	0.454**	0.408*	0.282†	0.170
	—	(0.166)	(0.196)	(0.157)	(0.214)
Education	−0.115**		−0.142		0.148
	(0.0241)		(0.109)		(0.120)
Years in office	−0.0180**		0.0798**		0.0535*
	(0.00396)		(0.0179)		(0.0212)
Ideology	0.152**		0.0395		−0.527*
	(0.0412)		(0.184)		(0.212)
Office characteristics					
Policy forbidding racial profiling	−0.433**		−0.621*		−0.738**
	(0.0552)		(0.258)		(0.273)
% of Asian or Hispanic deputies	−0.475		−1.551		−3.997†
	(0.409)		(1.808)		(2.133)
Latino sheriff	−0.537†		−0.952		1.349
	(0.286)		(1.265)		(1.251)
Latino Sheriff × Border Area	−0.139		−2.496		—
	(0.462)		(2.038)		—
County characteristics					
Vote for Obama 2008	−0.899**		−1.565		−4.313**
	(0.249)		(1.112)		(1.234)
Log (population)	0.00531		0.0888		0.245
	(0.0340)		(0.150)		(0.161)
% Urban population	−0.0729		4.040**		−2.607
	(0.341)		(1.508)		(1.666)
% Latino	0.0136*		0.0237		0.0143
	(0.00556)		(0.0247)		(0.0253)
% of foreign-born naturalized	−0.000327		−0.0123†		−0.0221**
	(0.00156)		(0.00688)		(0.00784)
Δ in % unemployment 2009–2011	−0.0229		0.227*		0.303**
	(0.0206)		(0.0912)		(0.107)
Δ in % foreign born 2000–2010	−0.00606		0.0442		0.0845
	(0.0162)		(0.0716)		(0.0744)
Constant	36.46**	2.381**	−156.9**	−1.018**	−106.2*
	(7.917)	(0.101)	(35.68)	(0.0971)	(42.44)
n	521	556	521	552	515
R^2/Pseudo R^2	.295	.013	.142	.005	.143

Standard errors are in parentheses. Immigration attitudes and number of immigration checks model use OLS regression. Policy on immigration models uses logistic regression. There are also issues with correlation between the Hispanic sheriff variable and the border area variable; as such, we avoid issues of multicollinearity, we use two generated variables of *Hispanic Sheriff × Border Area* and *White Sheriff × Non-Border Area* and include the Hispanic sheriff variable, thus making the comparison with white sheriffs in border areas. In some of the models, the interaction variables will not specify due to size and are not included; unfortunately, we cannot remedy this, as we simply do not have the necessary responses from Hispanic sheriffs in non-border areas. OLS = ordinary least squares.
†$p < .10$. *$p < .05$. **$p < .01$.

2. Independent Variables

There are also groups of words on the left side, each in its own row (ex: “Education”, “Years in office”, and all the way down, until “Constant”, which I’ll explain later). These are independent variables. Dependent variables, which we discussed before, are what the researchers are trying to explain, and independent variables are all the things the researchers think might explain them. Each model has one dependent variable, but can have lots and lots of independent variables. You can think of independent variables as causes, and dependent variables as the effect or outcome. Any outcome can and usually does have multiple causes. In the table above, Farris and Holman kindly group their independent variables to make it easy for the reader to understand: some of the causes are characteristics of the sheriffs in question, some are characteristics of the office, and some are characteristics of the county.

3. Coefficients

Each row has an independent variable, and then a bunch of numbers, one for each dependent variable. The numbers that are not in parentheses are called *coefficients*. A coefficient is the effect of the independent variable on the dependent variable after we "hold constant" or "control for" all the other variables with coefficients in that column. That just means that we're using statistical tools to make sure that we're actually picking up on the effect of the specific variable, and not the others. It's like saying 'even if this observation is *exactly the same* in all the other ways, here's the effect of this independent variable on the dependent variable'. In the table above, in the row for Education and the column for Immigration Attitudes Factor, the coefficient is -0.115. This means that the effect of education on Immigration Attitudes Factor is -0.115, even when we account for years in office, ideology, policies forbidding racial profiling, and so on. So even if we had sheriffs with exactly the same number of years in office, with exactly the same ideology, in counties with exactly the same policies forbidding racial profiling, the effect of education would be -0.115.

How to interpret that number -- what does it mean for an effect to be -0.115? -- depends on the statistical model the researcher is using[2], but the most important thing is the same for pretty much every model you'll ever encounter: the sign of the coefficient. The sign -- whether the number is positive or negative -- tells you whether the independent variable has a positive or negative effect on the dependent variable. In the above example, -0.115 is a negative number, meaning that as education (independent variable) increases, immigration attitudes factor (dependent variable) goes down (when the other variables are the same). And vice versa. As education goes down, immigration attitudes factor goes up. How much it goes up can be hard to tell in most models, as can whether an effect is big or small, but you can always tell whether the effect is positive or negative. And for a lot of papers, that's really all we care about, ultimately: does the independent variable increase or decrease the dependent variable? And is that still true even if we account for other possible explanations?

4. Constant

The constant -- sometimes listed as the intercept -- and often either listed in the first row or one of the last rows, is a particular kind of coefficient. It isn't associated with any variable, because it's the "effect" when all the independent variables are equal to zero. This is almost never something we actually care about, and a lot of the time it's not something that can ever exist, because not all of the variables *can be* zero. (Imagine a situation where you have GDP or population as an independent variable, for instance -- zero GDP or zero population would be weird.) Most of the time you can safely ignore the constant/intercept when interpreting a regression table, but never ignore it if you're actually running a model yourself.

---

[2] Ordinary least squares (OLS) regression -- also sometimes called linear regression or the linear model -- is a very common statistical model with the most straightforward interpretation. If you're reading an OLS regression (and it should say so in the title or caption of the table), that -0.115 means that a *one-unit increase* in education corresponds with a drop of 0.115 in Immigration Attitudes Factor.

5. **Standard Errors and Stars**

The other number associated with each pair of independent and dependent variables is the one that's in parentheses. This is called a "standard error". A standard error tell us how sure we are of the coefficient estimate. Because in statistics we're never totally sure -- the world is a noisy, uncertain place! -- all of our coefficients are really estimates, and our estimates are never absolutely right. Because we know that, since that's how statistics is, we produce this other number that tells us just how sure we are.

You don't really need to know what a standard error means.[3] What matters is how big it is compared with the coefficient it's associated with. If it's bigger than the coefficient or even roughly the same size, it means we're not at all sure of that estimate, and you can safely ignore it -- it might even be zero or the opposite sign. But if it's a lot smaller than the coefficient, that tells you we are pretty sure that the actual effect is the coefficient we estimated.

But you don't even really have to compare those numbers, because most tables also include symbols -- usually stars (*), but the table above also uses a small cross -- that give you the same information. Each table will have a key (often at the lower left or lower right), but as a rule of thumb: the more stars, the more sure we are of the estimate. If there's no symbol at all, we can't say that we have any certainty about the estimate.[4] If there are any stars, or lots of stars, you can conclude we are either pretty sure or very sure about the estimate.

If you combine this with the sign of the coefficient, you can make very important statements: we are fairly certain the independent variable has a positive (or negative, depending on the sign) effect on the dependent variable.

6. **Some Quick FAQs**

Why do some cells have "--" instead of a number or no number at all?
This means they aren't included as independent variables for that particular dependent variable.

What's n?
The number of observations. An observation is whatever the data are collected about. In the above table, each observation is a sheriff, so n is the number of sheriffs there is data about for each dependent variable. In other models, observations are individuals, countries, and so on.

What's R2 or Pseudo-R2?
This tells you how much of the variation in the dependent variable can be explained by the independent variables in the model. In short, if it's close to 1, there's not a lot left to explain. If it's close to 0, there's a lot more to explain. No value of this is good or bad -- a model that explains a lot is not necessarily better than a model that doesn't -- but it is descriptive.

Citations

Farris, Emily M., and Mirya R. Holman. "All politics is local? County sheriffs and localized policies of immigration enforcement." *Political Research Quarterly* 70, no. 1 (2017): 142-154.

---

[3] It's the standard deviation of the sampling distribution, if anyone's wondering. If you know what that means, you should probably be reading a more technical guide.
[4] There are debates about how sure we should be before we include stars, but it's always the case that more stars means more certainty.