Engaging Political Science Research

Reading Journal Articles: Argument-and-Evidence

You'll find that articles are usually filled with technical jargon. You should engage these details as best you can, but keep in mind the big picture.

1. What is the author's main **argument**? This usually comes as (1) one or more simple causal claims or (2) a more detailed theoretical model.
2. What is the author's **evidence** for their claim? Usually, this argument comes in the following form:
    1. In these data, higher [lower] values of X are associated with higher values of Y.
    2. This pattern could not be reasonably explained as spurious or reverse causation.
    3. The strength of this association could not be reasonably explained as due to chance.

Evidence: Sign-and-Significance

Linear Regression

aka least squares regression, ordinary least squares (OLS)

Earlier in the semester, we learned a simple method, that we called "the regression method." Using this approach, we model the average value of the outcome variable y as a linear function of the key explanatory variable x. Then we learned that we can include other explanatory variables and saw the standard multivariate regression model (in econometrics-style notation), so that

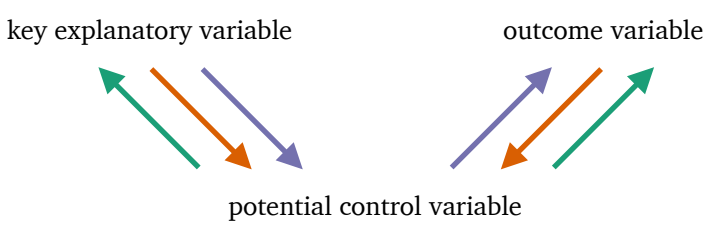
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + u.$$

In this model, y represents the outcome variable, and the x_i represents an explanatory variable. For simplicity, we can let x_1 represent the key explanatory variables and refer to all other x_i as "control variables."

Here's the key idea: **If we include the correct control variables, then β_1 should be about the causal effect of a one unit increase in x_1.**

But what control variables must we include and exclude? Here's a guide using causal arrows. The difference colors show several common causal relationship among the outcome, key explanatory variables.

Situation	Description	Include the potential control?
Confounding	The potential control causes both the key explanatory variable and the outcome variable.	YES! You must include all confounders!
Colliding	The potential control is caused by both the key explanatory variable and the outcome variable.	NO! You must not include any colliders!
Mediating	The key explanatory variable causes the potential control, and the potential control causes the outcome variable.	NO! You must not include any mediators!



Note that some researchers include non-linear terms in the model. These include polynomial terms, like x_1^2, and product terms, like $x_1 x_2$. If the key explanatory variable is included non-linearly (in a polynomial or product term), then you cannot easily use the guidelines below.

Other Regression Models

There are other "regression" methods that fit data slightly differently—that don't minimize the r.m.s. of the errors. These methods often lose the direct interpretation of the (slope) coefficients. Common examples include "logistic regression," "probit," "ordered probit," "negative binomial regression," and a "Cox proportional hazards model."

But the interpretation of the slope coefficients isn't as easy for these other models.

Interpretation for Linear Regression: A one-unit increase in x_1 leads to an increase of about β_1 in y, on average.

Interpretation for Many Other Types of Regression: An increase in x_1 leads to an increase [decrease] in y, on average, if β_1 is positive [negative].

Regression Tables

Political scientists usually report regression results in tables that include the coefficient (i.e., slope) estimates and other quantities. The table annotation should help you sort through the information.

To interpret this, I recommend the sign-and-significance method. Note that this does not include any information about the magnitude or size of the effect, which is a crucial part of the evidence. But it's complicated to extract that information, so you'll need to count on the author.

The Sign-and-Significance Method

1. Check that the key explanatory variable is included linearly. Make sure that it's not (1) multiplied by another variable or (2) squared, cubed, etc. If the key explanatory variable isn't included linearly, we need more complicated tools for interpretation.
2. Next, see if a coefficient of interest is statistically significant. Most tables report confidence intervals, *p*-values, standard error, or stars. We have some idea how this works, but you can use the following rules. These rules should be roughly equivalent and most tables provide starts.
   i. one or more stars
   ii. $p < .05$
   iii. SE < estimate/2
   iv. 95% confidence interval does not contain zero.
   v. Abs(t) > about 2
3. If the coefficient is significant, move to the next step. If the coefficient is not significant, we cannot be confident about the direction of the effect.
4. If the coefficient is significant, we can be confident that the variable has either a positive or negative relationship with the outcome. To know which, look at the sign of the coefficient. A positive sign indicates a positive relationship. A negative sign indicates a negative relationship.

Engage

Here are the question to ask yourself while reading a paper:

1. What's their claim, hypothesis, or theory?
2. Can the pattern in the data reasonably be explained by chance? If not, is it positive or negative?
3. Does the author control for <u>all</u> confounders? Avoid controlling for colliders and mediators?