What Do We Know About British Attitudes Toward Immigration? A Pedagogical Exercise of Regression

Steven V. Miller

Department of Political Science

CLEMSON UNIVERSITY

Goal for Today

1. Learn about immigration sentiment in the UK.
2. Teach students how to evaluate a regression table.

Check My Blog at svmiller.com/blog

What Do We Know About British Attitudes Toward Immigration? A Pedagogical Exercise of Sample Inference and Regression

Posted on March 23, 2020 by steve in R Political Science Teaching

This is a companion blog post to a presentation I was invited to give to some politics students in the United Kingdom, though this in-person presentation was unfortunately canceled in light of the COVID-19 pandemic.

What follows should not be interpreted as exhaustive of all the covariates of anti-immigration sentiment in the United Kingdom, or more generally. It clearly is not. Instead, the purpose of this presentation is to introduce these students to a quantitative approach to a social scientific problem in only 15 minutes and assuming no background knowledge on quantitative methods for the intended audience. As such, consider it an update to one of the most widely read pieces on my blog on how students should think about evaluating a regression table. It will ideally improve upon that, but I'll leave that determination if it does to the reader.

Immigration has overtaken the NHS as the most commonly mentioned worry of the British voter, according to an Ipsos Mori poll. (Photo: REX, via The Telegraph)

The post will also include some R code necessary to generate these results. There is only so much I can do within the allocated time to introduce students to a quantitative approach to social science. I don't get the opportunity to show them R code, though I would love if space and time permitted it. Toward that end, I will reference how I'm doing this in R with some code chunks in the post. The `_source` directory on the Github directory for my site will have the full code for this post. The particular source file is here.

Here are all the R packages that I'll use for the important stuff in this post.

A Regression Roadmap

Regression as we use it is a combination of hypothesis-testing and story-telling.

1. Know the bigger picture/puzzle.
2. Know the data.
3. Understand what the regression table is saying.
4. Understand what the regression table *isn't* saying.

Know the Bigger Picture/Puzzle

How positively do British people regard immigration/immigrants?

Why is this important? How can we know?

Know the Data

Scan the research design section for the following information:

1. The (primary) source of the data
2. The **unit of analysis** (i.e. who/what is being studied)
3. The **dependent variable** (i.e. the variable to be explained)
4. The **independent variable(s)** (i.e. what we be believe explains the dependent variable)

Know the Data

1. The **data:** European Social Survey (2018) for the UK
2. The **unit of analysis:** the individual respondent in the survey
    - Note: I subset the analysis to just those who were born in the UK.

Know the Data

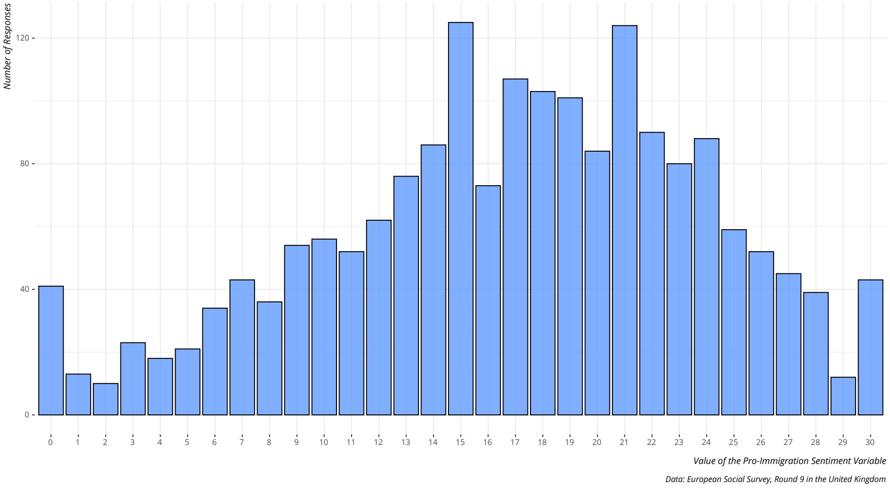
The **dependent variable** (*DV*) is an additive index [0:30] of three prompts:

- Is it generally bad or good for the UK's economy that immigrants come to live here?
  - (`imbgeco`) [0:10; bad:good]
- Is the UK's cultural life is generally undermined or enriched by immigrants?
  - (`imueclt`) [0:10; undermined:enriched]
- Is the UK made a worse or a better place to live by immigrants?
  - (`imwbcnt`) [0:10; worse:better]

Higher values = more pro-immigration sentiment.

A Bar Chart of Pro-Immigration Sentiment in the United Kingdom from the ESS Data (Round 9)

There's a natural heaping of 0s and 30s but the mean (16.891) approximates the median (17). I'd feel comfortable communicating exact differences on this scale.



Know the Data

The **independent variables** (*IVs*):

- *Age* (in years)
- *Education* (in years of education)
- *Gender* (1 if respondent is a woman)
- *Employment status* (1 if respondent is unemployed, but looking for work)
- *Household income* (in deciles)
- *Ideology* (on 11-point L-R scale)

Know the Data

Table 1: Descriptive Statistics for the Variables in Our Regression

Variable	**Mean**	**Std. Dev.**	**Median**	**Min.**	**Max.**	**N**
Immigration Sentiment	16.891	6.991	17	0	30	1850
Age	53.673	18.392	55	15	90	1893
Female	.541	.011	1	0	1	1905
Years of Education	14.049	3.630	13	3	41	1893
Unemployed	.019	.003	0	0	1	1905
Household Income (Deciles)	5.171	2.972	5	1	10	1615
Ideology (L to R)	4.96	1.945	5	0	10	1726

Understand What the Regression is Saying

(Linear/OLS) Regression is a tool for understanding a phenomenon of interest (immigration sentiment) as a linear function of some combination of predictors.

- Strong resemblance to the slope-intercept equation (`y = mx + b`)
- Flexible to include multiple predictors (i.e. **multiple regression**)

Understand What the Regression is Saying

We believe we can explain the *DV* for an individual (i) as:

$$
\begin{aligned}
\text{Immigration Sentiment} \quad = \quad & \beta_0 + \beta_1 * \text{Age} + \beta_2 * \text{Female} + \beta_3 * \text{Years of Education} + \\
& \beta_4 * \text{Unemployed} + \beta_5 * \text{Household Income} + \\
& \beta_6 * \text{Ideology} + \epsilon
\end{aligned}
$$

Of note:

- β_0: estimated immigration sentiment when all predictors are set to zero (i.e. *y*-intercept)
- $\beta_{1,2,3,4,5,6}$: estimated slopes of each *IV* on the *DV*
- ϵ: prediction errors, an unmeasured variable

Table 2: A Simple OLS Model of Pro-Immigration Sentiment in the United Kingdom

	Pro-Immigration Sentiment
Age	-0.002
	(0.010)
Female	-0.248
	(0.338)
Years of Education	0.488^{***}
	(0.049)
Unemployed	-1.102
	(1.204)
Household Income (Deciles)	0.338^{***}
	(0.061)
Ideology (L to R)	-0.583^{***}
	(0.088)
Constant	11.655^{***}
	(1.061)
N	1454

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

Unpacking the Regression Table

Take stock of three things from a regression table:

1. The numbers inside parentheses next to a variable.
2. The numbers *not* in parentheses next to a variable.
3. Some of those numbers not in parentheses have some asterisks next to them.

Let's start with the second item.

The Regression Coefficient

The number *not* inside a parentheses is a **regression coefficient**.

- These communicated the *estimated* change in the *DV* for a one-unit change in a particular *IV*.
- They are the betas (β) in the formula.
- The "Constant" is the *y*-intercept (β_0).

Use the coefficients to assess **negative** and **positive relationships**.

- **Positive:** as an *IV* increases, the *DV* also increases (and vice-versa).
- **Negative:** as an *IV* decreases, the *DV* increases (and vice-versa).

Table 3: A Simple OLS Model of Pro-Immigration Sentiment in the United Kingdom

	Pro-Immigration Sentiment
Age	-0.002
	(0.010)
Female	-0.248
	(0.338)
Years of Education	0.488^{***}
	(0.049)
Unemployed	-1.102
	(1.204)
Household Income (Deciles)	0.338^{***}
	(0.061)
Ideology (L to R)	-0.583^{***}
	(0.088)
Constant	11.655^{***}
	(1.061)
N	1454

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

The Standard Error

The number inside parentheses is the **standard error** of the regression coefficient.

- i.e. it's a measure of uncertainty around the estimate.

However, this is not an important quality by itself.

- It depends on its relationship with the regression coefficient.

The Asterisks

The asterisks are an indicator of **statistical significance**.

- Divide the regression coefficient over its standard error.
- This gives you a **t-statistic**.
- If the absolute value of the **t-statistic** is greater than 1.96 ("about 2"), you can feel confident rejecting a claim of zero.

tl;dr: what is the probability of us observing this coefficient and standard error if the true effect is zero?

- The asterisks are your visual cue to identify a "statistically significant" effect.

In our case, only education, income, and ideology have statistically significant effects.

Table 4: A Simple OLS Model of Pro-Immigration Sentiment in the United Kingdom

	Pro-Immigration Sentiment
Age	-0.002
	(0.010)
Female	-0.248
	(0.338)
Years of Education	0.488^{***}
	(0.049)
Unemployed	-1.102
	(1.204)
Household Income (Deciles)	0.338^{***}
	(0.061)
Ideology (L to R)	-0.583^{***}
	(0.088)
Constant	11.655^{***}
	(1.061)
N	1454

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

Understand What the Regression Table *Isn't* Saying

1. The "constant" (or y-intercept) is not a "variable."
   - It's just an estimate of y when everything else is zero.
2. The regression table doesn't test the regression model's multiple assumptions.
   - Look at your data and your model.
3. "Statistically significant is not itself 'significant.' "
   - "Significance" says nothing about the magnitude of the effect, only if you can discern it from zero.
4. More asterisks do not mean "more significance."
   - It won't tell you if one effect is greater than another, only that it's more precise.

Conclusion

Regression is both hypothesis-testing and story-telling. Some takeaways:

1. Make sure you're reading about the bigger picture/puzzle.
2. Take stock of the data.
   - (i.e. *DV*, *IVs*, unit of analysis, and data source).
3. Evaluate a regression table you read by the direction of the relationship (+ or -) and what effects are "significant."
4. Internalize what the regression table *isn't* telling you.
   - (i.e. magnitude effects, whether the model's assumptions are met).

See more information on my blog about this exercise: `http://svmiller.com/blog`

Table of Contents