Intermediate Quantitative Social Research

A Course Proposal

Steven V. Miller

Department of Political Science

CLEMSON UNIVERSITY

Goals for Today

1. Outline a course proposal for an intermediate quant methods class in social science research.
2. Introduce students to R (with examples)

About This Course

This is an intermediate course, assuming a previous course that covered:

- Operationalisation of research questions
- Hypothesis testing
- Basic mathematics for social scientists
- Research designs (e.g. experiments, surveys)
- Descriptive statistics
- OLS regression

What We Will Cover

1. The R programming language
2. Data visualization, descriptive statistics
3. Fitting/interpreting OLS/logistic models.
4. Fitting/interpreting mixed effects models.
5. Model diagnostics/comparisons.
6. Writing a quantitative research paper.

Why R, and How?

Why:

- High demand in private sector.
- R is free; everything else costs too much money.
- Rstudio is an excellent IDE (and also free).
- *Great* community support (e.g. StackOverflow, `#rstats` on Twitter).
- Unbeatable for data visualization and document prep (through R Markdown).

How:

- Every lecture comes with lab scripts.
- Everything goes on Github.
- Ample support on my course website/blog.
- You'll learn in part by mimicking.

See more at `http://svmiller.com/presentations`.

An Example: British Attitudes about Immigration/Immigrants

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.

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.

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)

Some Startup R Libraries We'll Need

```
library(tidyverse) # for all things workflow
library(stevedata) # for the data (ESS9GB)
library(stevemisc) # helper functions from my toy package

# Let's use {tidyverse} to create another DV
# This will equal 1 if respondent thinks immigrants
#  mostly undermine UK culture.

ESS9GB %>%
  mutate(imuecltd = ifelse(imueclt < 5, 1, 0)) -> ESS9GB
```

What We Can Do in R

We can get summary statistics, by region...

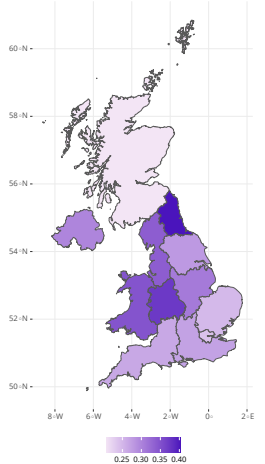
```
ESS9GB %>%
  group_by(region) %>%
  summarize(prop_undermine = mean(imuecltd, na.rm=T),
            mean_immigsent = mean(immigsent, na.rm=T))
```

```
## # A tibble: 12 x 3
##    region                    prop_undermine mean_immigsent
##    <chr>                              <dbl>          <dbl>
##  1 East Midlands (England)            0.310           16.4
##  2 East of England                    0.25            17.4
##  3 London                             0.241           18.0
##  4 North East (England)               0.404           14.7
##  5 North West (England)               0.339           15.5
##  6 Northern Ireland                   0.3             17.3
##  7 Scotland                           0.208           18.5
##  8 South East (England)               0.272           17.9
##  9 South West (England)               0.266           17.6
## 10 Wales                              0.348           15.8
## 11 West Midlands (England)            0.374           15.6
## 12 Yorkshire and the Humber           0.280           16.6
```

We Can Even Visualize These Findings With a Map

Percentage of Respondents Thinking Immigrants Undermine Culture, by Region

The sentiment is highest in North East (40%) and lowest in London (24%) and Scotland (20%).



Data: ?ESS9GB in (stevedata), by way of the European Social Survey (2018).

We Can Run a Few Regression Models

```
# Linear model
M1 <- lm(immigsent ~ agea + female + eduyrs + uempla + hinctnta +
           lrscale, data=ESS9GB)
# Logistic model
M2 <- glm(imuecltd ~ agea + female + eduyrs + uempla + hinctnta +
            lrscale, data=ESS9GB, family = binomial(link="logit"))
```

We Can Even Generate Fancy Regression Tables (in {modelsummary})

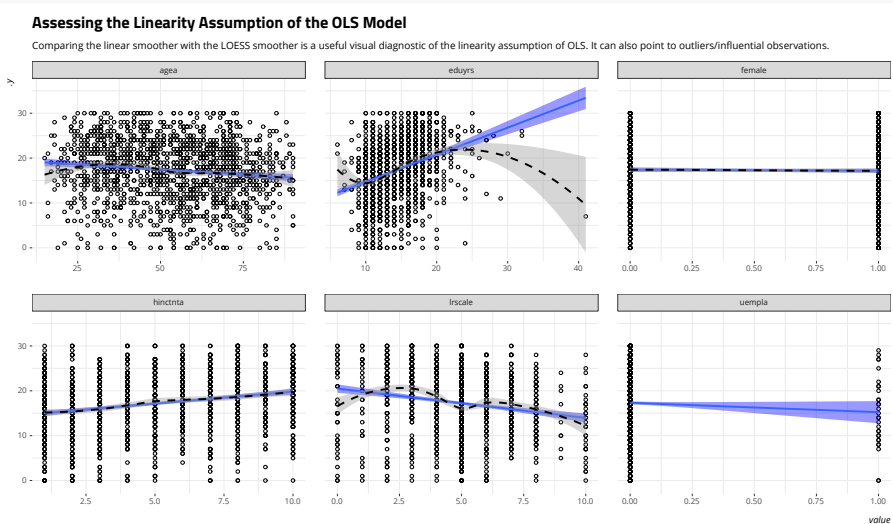
Table 1: Simple Models of Immigration Attitudes in the United Kingdom

	Pro-Immigration Sentiment	Immigrants Undermine Culture
Age	-0.002	0.003
	(0.010)	(0.004)
Female	-0.248	-0.130
	(0.338)	(0.122)
Years of Education	**0.488***	**-0.110***
	(0.049)	**(0.020)**
Unemployed	-1.102	0.398
	(1.204)	(0.396)
Household Income (Deciles)	**0.338***	**-0.087***
	(0.061)	**(0.023)**
Ideology (L to R)	**-0.583***	**0.120***
	(0.088)	**(0.032)**
Intercept	11.655*	0.303
	(1.061)	(0.398)
Num.Obs.	1454	1469

* $p < 0.05$

We Can Conveniently Do Some Model Diagnostics

```
linloess_plot(M1) # in {stevemisc}
```

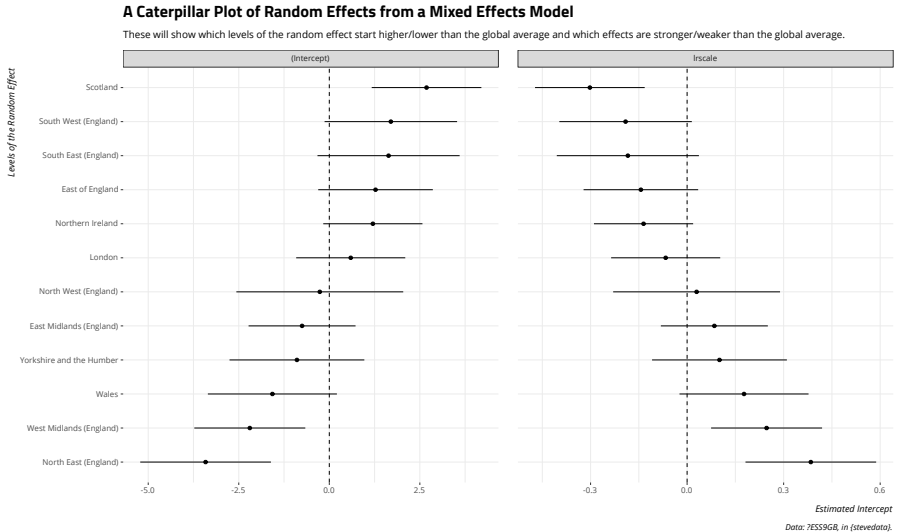


We Can Also Conveniently Run Mixed Effects Models

```
library(lme4) # everyone's go-to for mixed models

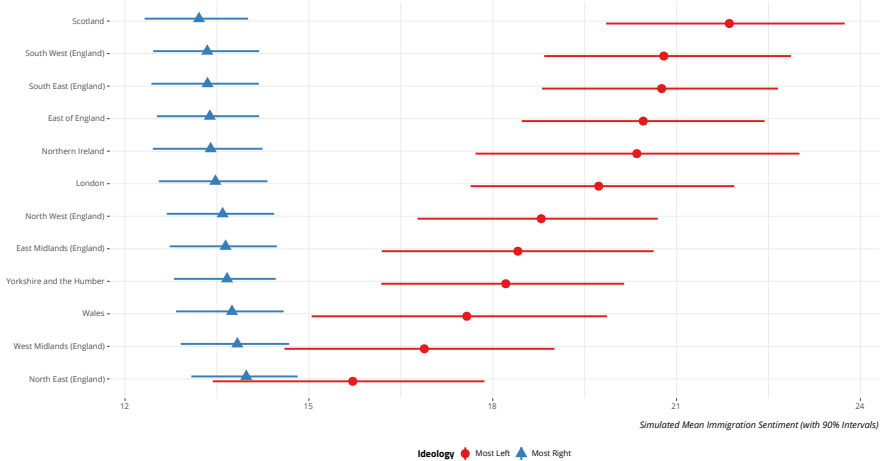
M3 <- lmer(immigsent ~ agea + female + eduyrs + uempla + hinctnta +
             lrscale + (1 + lrscale | region), data=ESS9GB)
M4 <- glmer(imuecltd ~ agea + female + eduyrs + uempla + hinctnta +
              lrscale + (1 + lrscale | region), data=ESS9GB,
            family = binomial(link="logit"))
#^ will want you to rescale your variables, and we'll talk about why you should do this.
```

```
show_ranef(M3, "region", reorder=FALSE) # in {stevemisc}
```



Post-Estimation Simulation of Mixed Models Will Tell You More About What Your Effects 'Look Like'

They suggest that the most left North East respondents aren't that different from the most right in their sentiment toward immigration. In Scotland: they're very different.



Data: ?ESS9GB, in {stevedata). Method: Simulation by multivariate normal distribution of coefficients and variance-covariance matrix.

We Can Also Write Our Reports in R (R Markdown)

```
---
output: stevetemplates::word
title: "My Analysis on Attitudes About Immigration in the United Kingdom"
author: A. Awesome Student
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE, message=FALSE,
                      dpi = 600,
                      warning=FALSE,
                      fig.width = 8.5,
                      fig.path = "paper-example/figs/",
                      cache.path = "paper-example/_cache/")

library(tidyverse) # for all things workflow
library(stevedata) # for the data
library(stevemisc) # graph formatting, other helper functions
library(lme4) # everyone's go-to for mixed models
library(modelsummary) # regression tables

ESS9GB %>%
  mutate(imuecltd = ifelse(imueclt < 5, 1, 0)) -> ESS9GB

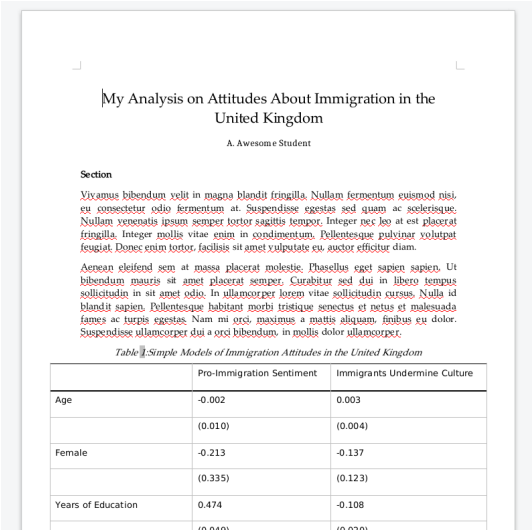
M3 <- lmer(immigsent ~ agea + female + eduyrs + uempla + hinctnta +
             lrscale + (1 + lrscale | region), data=ESS9GB)
M4 <- glmer(imuecltd ~ agea + female + eduyrs + uempla + hinctnta +
              lrscale + (1 + lrscale | region), data=ESS9GB,
            family = binomial(link="logit"))

```

Section
```

This entire presentation was written in R/R Markdown.

The Ouput from my Word Template



My Analysis on Attitudes About Immigration in the United Kingdom

A. Awesome Student

Section

Vivamus bibendum velit in magna blandit fringilla. Nullam fermentum euismod nisi, eu consectetur odio fermentum at. Suspendisse egestas sed quam ac scelerisque. Nullam venenatis ipsum semper tortor sagittis tempor. Integer nec leo at est placerat fringilla. Integer mollis vitae enim in condimentum. Pellentesque pulvinar volutpat feugiat. Donec enim tortor, facilisis sit amet vulputate eu, auctor efficitur diam.

Aenean eleifend sem at massa placerat molestie. Phasellus eget sapien sapien. Ut bibendum mauris sit amet placerat semper. Curabitur sed dui in libero tempus sollicitudin in sit amet odio. In ullamcorper lorem vitae sollicitudin cursus. Nulla id blandit sapien. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nam mi orci, maximus a mattis aliquam, finibus eu dolor. Suspendisse ullamcorper dui a orci bibendum, in mollis dolor ullamcorper.

Table 1: Simple Models of Immigration Attitudes in the United Kingdom

	Pro-Immigration Sentiment	Immigrants Undermine Culture
Age	-0.002	0.003
	(0.010)	(0.004)
Female	-0.213	-0.137
	(0.335)	(0.123)
Years of Education	0.474	-0.108
	(0.040)	(0.020)

Conclusion

This applied course would teach students many real-world skills.

- Statistical concepts (e.g. logistic regression, mixed models)
- Applied methodological skills (all in R)
- Reproducibility/workflow techniques (all in R, with help from my suite of R packages)

It would also teach/do more than I can cover in this presentation.

- e.g. diagnostics, theory, and other good practices

See my website (`http://svmiller.com`) for more.

Table of Contents