Assignment 2 - Solutions

Mark Deming

Due 2/26

```r
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE)

# Load packages here
require(tidyverse)    # For data manipulation
```

```
## Loading required package: tidyverse
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
require(readr)        # For reading in data
require(haven)        # For reading in Stata data
```

```
## Loading required package: haven
```

```r
require(here)         # For setting working directory
```

```
## Loading required package: here
## here() starts at /Users/mdeming/Library/CloudStorage/Box-Box/teaching/courses/independent_study
```

```r
require(estimatr)     # For robust standard errors
```

```
## Loading required package: estimatr
```

```r
require(modelsummary) # For nice regression tables
```

```
## Loading required package: modelsummary
## ‘modelsummary‘ 2.0.0 now uses ‘tinytable‘ as its default table-drawing
##   backend. Learn more at: https://vincentarelbundock.github.io/tinytable/
##
```

```
## Revert to ‘kableExtra‘ for one session:
##
##   options(modelsummary_factory_default = ’kableExtra’)
##   options(modelsummary_factory_latex = ’kableExtra’)
##   options(modelsummary_factory_html = ’kableExtra’)
##
## Silence this message forever:
##
##   config_modelsummary(startup_message = FALSE)
```

```
require(plm)          # For panel data analysis
```

```
## Loading required package: plm
##
## Attaching package: ’plm’
##
## The following objects are masked from ’package:dplyr’:
##
##     between, lag, lead
```

```
# Load data here
bes05 <- read_dta(here("datasets","koch-nicholson_2016","bes05_short.dta"))
bes10 <- read_dta(here("datasets","koch-nicholson_2016","bes10_short.dta"))
casualties <- read_delim(here("datasets","koch-nicholson_2016","ukregion_cas.tab"), delim = "\t")
```

```
## Rows: 22 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## dbl (3): region, year, region_cas
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
districtdata <- read_delim(here("datasets","koch-nicholson_2016","0501districtdata.tab"), delim = "\t")
```

```
## Rows: 22 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (1): area
## dbl (6): population, income_pc, year, unemploy_rate, pctwhite, region
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

Overview

You will do a complete replication of Table 4 from Koch and Nicholson’s article (2016), “Death and Turnout”. This is part 2 of 2.

You will require the four datasets that you used in Assignment 1. All datasets are at Deming’s GitHub page: HERE:

- bes05_short.dta
- bes10_short.dta
- ukregion_cas.tab
- 0501districtdata.tab

Throughout, you should use `dplyr` functions and syntax whenever possible.

Get started

1. Load the packages below. You may need to install them first.

- `tidyverse` (contains `dplyr` and `ggplot2`)
- `readr` (for importing `.tab` formatted data)
- `haven` (for importing `.dta` formatted data)
- `here` (recommended but not required. You might read about how `here()` works.)
- `estimatr` (for robust standard errors)
- `modelsummary` (for nice regression tables)
- `plm` (for panel data analysis)

2. Import the four datasets above in the `setup` chunk.

3. In the chunk below, clean, append, and merge the datasets as you did in Assignment 1 (up through the "Explore" section). See if you can clean `bes05` and `bes10` in two long strings of piped code.

```
# 1. Clean bes05 in a single string of piped code
bes05 <- bes05 %>%
  rename(region = pre_q1,
         labor_iraq = pre_q13,
         conserve_iraq = pre_q23,
         partyid = pre_q29,
         party_strength = pre_q33,
         likelyvote = pre_q34,
         blair_competent = pre_q50,
         executive_approval = pre_q68,
         gov_party_approve = pre_q84,
         perception_economy = pre_q92,
         attention = pre_q141,
         birthyr = pre_q148,
         education = pre_q156,
         income = pre_q163,
         race = pre_q174,
         gender = pre_q180,
         marital_status = pre_q158,
         british_iraq = pre_q128,
         weights = pre_w8) %>%
  mutate(party_strength = if_else(party_strength == 4, NA_real_, party_strength),
         labor_iraq = if_else(labor_iraq == 6, NA_real_, labor_iraq),
         perception_economy = ifelse(perception_economy == 6, NA_real_, perception_economy),
         likelyvote = ifelse(likelyvote == 12, NA_real_, likelyvote)) %>%
  rename(pmtherm = executive_approval,
         pmwar = labor_iraq) %>%
  mutate(year = 2005,
         age = 2005 - birthyr)
```

```r
# 2. Clean bes10 in a single string of piped code
bes10 <- bes10 %>%
  rename(region = aaq1,
         labor_afghan = aaq13,
         conserve_afghan = aaq22,
         partyid = aaq28,
         party_strength = aaq32,
         likelyvote = aaq33,
         brown_competent = aaq81,
         executive_approval = aaq52,
         gov_party_approve = aaq63,
         perception_economy = aaq87,
         attention = aaq131,
         birthyr = aaq151,
         education = aaq159,
         income = aaq166,
         race = aaq177,
         gender = aaq186,
         marital_status = aaq161,
         british_afghan = aaq116,
         weights = w8_f) %>%
  mutate(party_strength = if_else(party_strength == 4, NA_real_, party_strength),
         labor_afghan = ifelse(labor_afghan == 6, NA_real_, labor_afghan),
         perception_economy = ifelse(perception_economy == 6, NA_real_, perception_economy),
         likelyvote = ifelse(likelyvote == 12, NA_real_, likelyvote),
         income = ifelse(income == 17, NA_real_, income)) %>%
  rename(pmtherm = executive_approval,
         pmwar = labor_afghan) %>%
  mutate(year = 2010,
         age = birthyr)

# 3. Append bes05 and bes10
bes0510 <- bind_rows(bes05, bes10)

# 4. Merge the appended bes dataset with casualties
bes0510casmerge <- bes0510 %>%
  left_join(casualties, by = c("region","year"))

# 5. Merge the bes-casualties dataset with the district data. Then, create the female,
# low_attention, married, and partstrength variables. Do this in a single string
# of piped code.
bes_final_data <- bes0510casmerge %>%
  left_join(districtdata, by = c("region","year")) %>%
  mutate(white = ifelse(race == 1, 1, 0),
         female = ifelse(gender == 2, 1, 0),
         low_attention = ifelse(attention < 4, 1, 0),
         married = ifelse(marital_status == 1, 1, 0),
         partstrength = ifelse(party_strength == 1, 1, 0))
```

Ordinary least squares model

4. Use `lm_robust()` to write a regression model that approximates Table 4 from Koch and Nicholson (2016). Do not include region dummies in the model.

```
ols_mod <- lm_robust(likelyvote ~
                     region_cas +
                     as.factor(low_attention) +
                     as.factor(low_attention) * region_cas +
                     birthyr +
                     partstrength +
                     female +
                     white +
                     married +
                     income +
                     education +
                     perception_economy +
                     pmtherm +
                     pmwar +
                     income_pc +
                     unemploy_rate +
                     pctwhite,
                     data = bes_final_data)
```

5. Use `modelsummary()` to display your results above in a nice regression table. Add informative coefficient labels following Table 4.

```
# Map variable labels to coefficients
coef_labels <- c(
                "region_cas" = "Local Casualties",
                "as.factor(low_attention)1" = "Low Attention",
                "region_cas:as.factor(low_attention)1" = "Attention x Casualties",
                "female" = "Female",
                "married" = "Married",
                "income" = "Income Level",
                "education" = "Education",
                "birthyr" = "Year Born",
                "white" = "White",
                "partstrength" = "Partisan Strength",
                "perception_economy" = "Perception of the Economy",
                "pmtherm" = "Executive Approval",
                "pmwar" = "War Approval",
                "income_pc" = "Median Incom",
                "unemploy_rate" = "Unemployment Rate",
                "pctwhite" = "% White",
                "(Intercept)" = "Constant")

# Display regression results
modelsummary(ols_mod,
             title = "Table 4: The Effect of Local Casualties on Voting in 2005 and 2010 U.K. Elections
             output = "markdown",
             stars = TRUE,
             coef_map = coef_labels,
             gof_omit = "IC|Log|RMSE")
```

6. Write a short paragraph of interpretation of the regression results above. Focus on the main coefficients: `region_cas`, `low_attention`, and the interaction term `low_attention:region_cas`. (What does the interaction term denote?)

Random effects model

The `bes` data are panel data: the same respondents are surveyed in 2005 and 2010. We can use a random effects model to account for the panel structure.

7. Create a new dataset that is a duplicate the Koch and Nicholson data above (2016). Designate the dataset as panel data using an appropriate function from the `plm` package. Use the `besid` and `year` variables as the index.

```
bes_panel <- pdata.frame(bes_final_data, index = c("besid", "year"))
```

8. Use `plm()` to write a random effects model that closely approximates Table 4 from Koch and Nicholson (2016). You should include region dummies in the model.

```
plm_mod <- plm(
            likelyvote ~
            region_cas +
            as.factor(low_attention) +
            as.factor(low_attention) * region_cas +
            birthyr +
            partstrength +
            female +
            white +
            married +
            income +
            education +
            perception_economy +
            pmtherm +
            pmwar +
            income_pc +
            unemploy_rate +
            pctwhite +
            as.factor(region),
  data = bes_panel,
  model = "random"
)

coef_labels <- c(
                 "region_cas" = "Local Casualties",
                 "as.factor(low_attention)1" = "Low Attention",
                 "region_cas:as.factor(low_attention)1" = "Attention x Casualties",
                 "female" = "Female",
                 "marriage" = "Married",
                 "income" = "Income Level",
                 "education" = "Education",
                 "birthyr" = "Year Born",
                 "white" = "White",
                 "partstrength" = "Partisan Strength",
                 "perception_economy" = "Perception of the Economy",
```

Table 1: Table 4: The Effect of Local Casualties on Voting in 2005 and 2010 U.K. Elections (by District)

	(1)
Local Casualties	0.002
	(0.003)
Low Attention	-2.111***
	(0.225)
Attention x Casualties	0.022**
	(0.007)
Female	0.141*
	(0.061)
Married	0.208**
	(0.066)
Income Level	0.044***
	(0.009)
Education	-0.012*
	(0.006)
Year Born	-0.022***
	(0.002)
White	0.565**
	(0.216)
Partisan Strength	0.557***
	(0.063)
Perception of the Economy	0.028
	(0.035)
Executive Approval	0.012
	(0.013)
War Approval	-0.015
	(0.031)
Median Incom	0.000+
	(0.000)
Unemployment Rate	0.048
	(0.031)
	(0.605)
Constant	7.886***
	(1.152)
Num.Obs.	4847
R2	0.125
R2 Adj.	0.122

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```r
                 "pmtherm" = "Executive Approval",
                 "pmwar" = "War Approval",
                 "income_pc" = "Median Incom",
                 "unemploy_rate" = "Unemployment Rate",
                 "pctwhite" = "% White",
                 "(Intercept)" = "Constant"
)

modelsummary(plm_mod,
             output = "markdown",
             stars = TRUE,
             gof_omit = "IC|Log|RMSE",
             ceof_map = coef_labels,
             vcov = "HC0",  # Clustered SEs
             coef_map = coef_labels)
```

9. Write a short paragraph of interpretation of the regression results above. In addition to interpreting the main coefficients, write 2-3 sentences about the random effects model. What does the random effects model add to the OLS model?

Wrap Up

When you finish:

- Knit this RMD to PDF.
- Review the RMD for completeness, accuracy, and neatness.
- Submit the RMD and PDF.

	(1)
Local Casualties	0.001
	(0.004)
Low Attention	-2.046***
	(0.143)
Attention x Casualties	0.022***
	(0.005)
Female	0.150*
	(0.066)
Income Level	0.046***
	(0.009)
Education	-0.011+
	(0.006)
Year Born	-0.023***
	(0.003)
White	0.536**
	(0.181)
Partisan Strength	0.476***
	(0.081)
Perception of the Economy	0.014
	(0.032)
Executive Approval	0.022+
	(0.012)
War Approval	-0.012
	(0.030)
Median Incom	0.000
	(0.000)
Unemployment Rate	0.087+
	(0.045)
	(1.263)
Constant	12.032***
	(2.823)
Num.Obs.	4847
R2	0.194
R2 Adj.	0.190
Std.Errors	HC0

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$