

Preliminary Analysis of ANES Data

Patrick

2025-03-19

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>. Hello this is nice. We can write some code. Add some comments.

$$\bar{x} = 1/N \sum x_i$$

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
# this is a usual comment
```

```
1 + 1
```

```
[1] 2
```

```
2 + 4
```

```
[1] 6
```

You can add options to executable code like this

```
[1] 4
```

The `echo: false` option disables the printing of code (only output is displayed).

Load Packages and Data

First we load our packages. We'll add more later.

```
library(here)
library(tidyverse)
library(modelsummary)
library(marginaleffects)
```

Now let's load our data. We will use the script that we have already written.

```
source(here("code/01-load_anes.R"))

anes
```

```
# A tibble: 8,280 x 13
   id mode    age female   edu college social misinfo_russia
   <dbl> <fct> <dbl>   <dbl> <dbl>   <dbl>   <dbl>         <dbl>
1 200015 Web     46     0     6     1 0.875           1
2 200022 Web     37     1     3     0 0.375          -1
3 200039 Web     40     1     2     0 0.5            -1
4 200046 Web     41     0     4     0 0.5            -1
5 200053 Web     72     0     8     1 0.125          -1
6 200060 Web     71     1     3     0 0             -1
7 200084 Web     37     1     4     0 0.375           1
8 200091 Web     45     1     2     0 0.25           -1
9 200107 Web     70     1     2     0 0.5            -1
10 200114 Web     43     0     4     0 0.875           1
# i 8,270 more rows
# i 5 more variables: confident_russia <dbl>, misconf_russia <dbl>,
#   misinfo_warm <dbl>, confident_warm <dbl>, misconf_warm <dbl>
```

Presenting Regression Results

First we have to run a few regressions.

```
m1a <- lm(misconf_russia ~ social + college +
          age + female + mode, data = anes)
summary(m1a)
```

```
Call:
lm(formula = misconf_russia ~ social + college + age + female +
    mode, data = anes)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.9979	-0.4819	-0.1756	0.3961	1.6045

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0678304	0.0354383	1.914	0.0557 .
social	-0.1700124	0.0364281	-4.667	3.11e-06 ***
college	-0.2116326	0.0145857	-14.510	< 2e-16 ***
age	-0.0044224	0.0004877	-9.069	< 2e-16 ***
female	0.0140873	0.0144537	0.975	0.3298
modePhone	0.0490508	0.0582347	0.842	0.3997
modeVideo	-0.0784512	0.0373307	-2.102	0.0356 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5963 on 6885 degrees of freedom

(1388 observations deleted due to missingness)

Multiple R-squared: 0.04497, Adjusted R-squared: 0.04414

F-statistic: 54.03 on 6 and 6885 DF, p-value: < 2.2e-16

```
m1b <- lm(misconf_russia ~ social + I(social^2) + college +
    age + female + mode, data = anes)
summary(m1b)
```

Call:

```
lm(formula = misconf_russia ~ social + I(social^2) + college +
    age + female + mode, data = anes)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.9723	-0.4790	-0.1787	0.3940	1.6012

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0446459	0.0371885	1.201	0.2300

social	0.0316291	0.1048123	0.302	0.7628
I(social^2)	-0.2498675	0.1217871	-2.052	0.0402 *
college	-0.2126240	0.0145903	-14.573	<2e-16 ***
age	-0.0044698	0.0004881	-9.158	<2e-16 ***
female	0.0125818	0.0144690	0.870	0.3846
modePhone	0.0528310	0.0582503	0.907	0.3645
modeVideo	-0.0805521	0.0373360	-2.157	0.0310 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5962 on 6884 degrees of freedom
(1388 observations deleted due to missingness)

Multiple R-squared: 0.04555, Adjusted R-squared: 0.04458

F-statistic: 46.94 on 7 and 6884 DF, p-value: < 2.2e-16

```
m1c <- lm(misconf_russia ~ social * college + I(social^2) * college +
          age + female + mode, data = anes)
summary(m1c)
```

Call:

```
lm(formula = misconf_russia ~ social * college + I(social^2) *
    college + age + female + mode, data = anes)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.9702	-0.4753	-0.1705	0.3958	1.6010

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0050335	0.0396223	0.127	0.89891
social	0.2731156	0.1378411	1.981	0.04759 *
college	-0.1077620	0.0390652	-2.759	0.00582 **
I(social^2)	-0.4985179	0.1658486	-3.006	0.00266 **
age	-0.0044948	0.0004881	-9.209	< 2e-16 ***
female	0.0137248	0.0144742	0.948	0.34305
modePhone	0.0563612	0.0582381	0.968	0.33319
modeVideo	-0.0799365	0.0373191	-2.142	0.03223 *
social:college	-0.5860933	0.2100605	-2.790	0.00528 **
college:I(social^2)	0.5910766	0.2450554	2.412	0.01589 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5959 on 6882 degrees of freedom
(1388 observations deleted due to missingness)
Multiple R-squared: 0.04673, Adjusted R-squared: 0.04548
F-statistic: 37.48 on 9 and 6882 DF, p-value: < 2.2e-16

Regression Tables

```
modelsummary(list(m1a, m1b, m1c),
  coef_map = c("social" = "Social Media Exposure",
    "college" = "College Education",
    "social:college" = "Social Media X College",
    "I(social^2)" = "Social Media Squared",
    "college:I(social^2)" = "Social Media Squared X College",
    "age" = "Age",
    "female" = "Gender (Female)",
    "modePhone" = "Survey Mode (Phone)",
    "modeVideo" = "Survey Mode (Video)",
    "(Intercept)" = "Constant"),
  gof_map = c("nobs", "r.squared", "rmse"),
  vcov = "HC3")
```

Code Appendix

```
# =====
# Title: Load and Recode ANES Data
# Author: Patrick
# Date: 12/03/25
# =====

# Packages -----

library(here)
library(tidyverse)
```

	(1)	(2)	(3)
Social Media Exposure	−0.170 (0.037)	0.032 (0.107)	0.273 (0.145)
College Education	−0.212 (0.015)	−0.213 (0.015)	−0.108 (0.040)
Social Media X College			−0.586 (0.213)
Social Media Squared		−0.250 (0.122)	−0.499 (0.173)
Social Media Squared X College			0.591 (0.245)
Age	−0.004 (0.000)	−0.004 (0.000)	−0.004 (0.000)
Gender (Female)	0.014 (0.015)	0.013 (0.015)	0.014 (0.015)
Survey Mode (Phone)	0.049 (0.066)	0.053 (0.066)	0.056 (0.066)
Survey Mode (Video)	−0.078 (0.035)	−0.081 (0.035)	−0.080 (0.035)
Constant	0.068 (0.036)	0.045 (0.038)	0.005 (0.041)
Num.Obs.	6892	6892	6892
R2	0.045	0.046	0.047
RMSE	0.60	0.60	0.60

```

# Load raw data -----

raw <- read_csv(here("data/anes_timeseries_2020_csv_20220210.csv"))

# There were some encoding issues but they don't affect variables we use
# problems(raw)
# colnames(raw)[1509]

# Data recoding -----

# table(raw$V202550, useNA = "always")

anes <- raw %>%
  mutate(
    across(starts_with("V202541"),
      ~ ifelse(.x >= 0, .x, NA)
    )
  ) %>%
  transmute(
    ## Survey meta info
    id = V200001,
    mode = recode_factor(V200002,
      `3` = "Web",
      `2` = "Phone",
      `1` = "Video"),

    ## Sociodemographics: age, gender, education
    age = na_if(V201507x, -9),
    female = na_if(V201600, -9) - 1,
    edu = recode(V201510,
      `-9` = NA_real_,
      `-8` = NA_real_,
      `95` = NA_real_),
    college = as.numeric(edu >= 6),

    ## Social media exposure
    social = rowSums(across(V202541a:V202541h)) / 8,
    # social = (V202541a + V202541b + V202541c + V202541d +
    #   V202541e + V202541f + V202541g + V202541h) / 8,
  )

```

```

## Misinformation index: high values = belief in conspiracies
misinfo_russia = recode(V202549,
  `1` = -1,
  `2` = 1,
  .default = NA_real_),
confident_russia = ifelse(V202550 > 0, (V202550 - 1)/4, NA),
misconf_russia = misinfo_russia * confident_russia,
misinfo_warm = recode(V202555,
  `1` = -1,
  `2` = 1,
  .default = NA_real_),
confident_warm = ifelse(V202556 > 0, (V202556 - 1)/4, NA),
misconf_warm = misinfo_warm * confident_warm
)

# table(raw$V202556, anes$confident_warm, useNA = "always")
# hist(anes$misconf_warm)

```