**In-class exercise (Instructors: email** [**favero@american.edu**](mailto:favero@american.edu) **for an answer key)**

use https://minusthemath.com/data/covid.dta

*Dataset description*

This dataset contains results of a survey of adult individuals in the US (recruited via the crowd-work survey platform Prolific).[[1]](#footnote-1) The survey was fielded on April 3, 2020—shortly after most US states had issued initial stay-at-home orders in response to the COVID-19 pandemic. The following variables are analyzed:

* gender: whether respondent identifies as male, female, non-binary, or prefers not to answer
* c\_age: respondent age, measured in years
* party: whether respondent’s preferred political party is Republican, Democratic, or other
* edu: the highest level of schooling that the respondent has completed
  + 1= Less than a high school diploma
  + 2= High school degree or equivalent
  + 3= Some college, no degree
  + 4= Associate degree
  + 5= Bachelor’s degree
  + 6= Master’s degree
  + 7= Professional degree
  + 8= Doctorate
* closebusiness: respondent agreement with the statement “The government should require all nonessential businesses in my area to close their on-site operations for at least the next 2 weeks”; measured on a 0-10 scale with 0 indicating strong disagreement and 10 indicating strong agreement
* maxweeks: anticipated social distancing duration, measured in weeks; respondents could enter any whole number; the survey prompt said “If officials advised it, I could see myself generally staying at home and avoiding social contact for up to \_\_\_ weeks”

**Q1**. Suppose I want to see if gender is associated with political party. What kind of analysis can I do? What do I find?

**Q2**. Is party related to anticipated social distancing duration? How can I test this? What do I find?

**Q3**. How about party and education level?

**Q4**. What about age and anticipated social distancing duration?

**Q5**. Belief that businesses should be closed and anticipated social distancing duration?

**Q6**. Examine the output in Table 8. First, write out the regression equation based on the output here.

**Q7**. Using the equation you just wrote out, tell me the predicted value of anticipated social distancing duration for a respondent who is 30 years old and has a 7 on the “close businesses” attitude question.

**Q8**. How about a 30 year old with a 10 on the “close businesses” attitude question? What is the predicted value for this hypothetical respondent?

**Q9**. Continue examining the output in Table 8 of the exam appendix. Focusing on the 3 S’s (significance, sign, and size), interpret the regression results for the two main independent variables.

**Table 1**

. tab gender party, row

+----------------+

| Key |

|----------------|

| frequency |

| row percentage |

+----------------+

| What is your preferred political

| party? - Selected Choice

What is your gender? | Republica Democrati Other | Total

---------------------+---------------------------------+----------

Prefer not to answer | 0 3 1 | 4

| 0.00 75.00 25.00 | 100.00

---------------------+---------------------------------+----------

Male | 179 436 137 | 752

| 23.80 57.98 18.22 | 100.00

---------------------+---------------------------------+----------

Female | 120 480 120 | 720

| 16.67 66.67 16.67 | 100.00

---------------------+---------------------------------+----------

Non-binary | 0 17 8 | 25

| 0.00 68.00 32.00 | 100.00

---------------------+---------------------------------+----------

Total | 299 936 266 | 1,501

| 19.92 62.36 17.72 | 100.00

**Table 2**

. bys party: sum maxweeks if party != 3, detail

--------------------------------------------------------------------------------

-> party = Republican

maxweeks

-------------------------------------------------------------

Percentiles Smallest

1% 1 0

5% 2 1

10% 3 1 Obs 300

25% 4 1 Sum of Wgt. 300

50% 8 Mean 17.65

Largest Std. Dev. 35.79532

75% 16 104

90% 51 156 Variance 1281.305

95% 52 200 Skewness 9.11839

99% 130 500 Kurtosis 114.2717

--------------------------------------------------------------------------------

-> party = Democratic

maxweeks

-------------------------------------------------------------

Percentiles Smallest

1% 2 0

5% 3 1

10% 4 1 Obs 936

25% 6 1 Sum of Wgt. 936

50% 12 Mean 1.07e+18

Largest Std. Dev. 3.27e+19

75% 20 156

90% 52 520 Variance 1.07e+39

95% 52 9999 Skewness 30.54507

99% 100 1.00e+21 Kurtosis 934.0011

**Table 3**

**. replace maxweeks = 104 if maxweeks > 104 // windsorize 12 extreme values**

(12 real changes made)

. bys party: sum maxweeks if party != 3, detail

--------------------------------------------------------------------------------

-> party = Republican

# of Weeks Willing to Isolate

-------------------------------------------------------------

Percentiles Smallest

1% 1 0

5% 2 1

10% 3 1 Obs 300

25% 4 1 Sum of Wgt. 300

50% 8 Mean 15.83667

Largest Std. Dev. 19.98653

75% 16 104

90% 51 104 Variance 399.4615

95% 52 104 Skewness 2.720826

99% 104 104 Kurtosis 10.82837

--------------------------------------------------------------------------------

-> party = Democratic

# of Weeks Willing to Isolate

-------------------------------------------------------------

Percentiles Smallest

1% 2 0

5% 3 1

10% 4 1 Obs 936

25% 6 1 Sum of Wgt. 936

50% 12 Mean 16.49573

Largest Std. Dev. 17.76754

75% 20 104

90% 52 104 Variance 315.6855

95% 52 104 Skewness 2.470056

99% 100 104 Kurtosis 10.09715

**Table 4**

. tab edu party, col

+-------------------+

| Key |

|-------------------|

| frequency |

| column percentage |

+-------------------+

What is the highest |

degree or level of | What is your preferred political

school you have | party? - Selected Choice

completed? | Republica Democrati Other | Total

----------------------+---------------------------------+----------

Less than a high scho | 2 8 4 | 14

| 0.67 0.85 1.50 | 0.93

----------------------+---------------------------------+----------

High school degree or | 34 95 29 | 158

| 11.37 10.15 10.90 | 10.53

----------------------+---------------------------------+----------

Some college, no degr | 58 237 71 | 366

| 19.40 25.32 26.69 | 24.38

----------------------+---------------------------------+----------

Associate degree | 30 73 30 | 133

| 10.03 7.80 11.28 | 8.86

----------------------+---------------------------------+----------

Bachelor’s degree | 115 357 93 | 565

| 38.46 38.14 34.96 | 37.64

----------------------+---------------------------------+----------

Master’s degree | 42 125 34 | 201

| 14.05 13.35 12.78 | 13.39

----------------------+---------------------------------+----------

Professional degree | 9 24 4 | 37

| 3.01 2.56 1.50 | 2.47

----------------------+---------------------------------+----------

Doctorate | 9 17 1 | 27

| 3.01 1.82 0.38 | 1.80

----------------------+---------------------------------+----------

Total | 299 936 266 | 1,501

| 100.00 100.00 100.00 | 100.00

**Table 5**

. bys party: sum edu

--------------------------------------------------------------------------------------

-> party = Republican

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

edu | 299 4.434783 1.492125 1 8

--------------------------------------------------------------------------------------

-> party = Democratic

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

edu | 936 4.316239 1.440586 1 8

--------------------------------------------------------------------------------------

-> party = Other

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

edu | 266 4.135338 1.372407 1 8

**Table 6**

. corr c\_age maxweeks

(obs=1,491)

| c\_age maxweeks

-------------+------------------

c\_age | 1.0000

maxweeks | 0.0285 1.0000

**Table 7**

. corr closebusiness maxweeks

(obs=1,502)

| closeb~s maxweeks

-------------+------------------

closebusin~s | 1.0000

maxweeks | 0.1355 1.0000

**Table 8**

. reg maxweeks c\_age closebusiness

Source | SS df MS Number of obs = 1,491

-------------+---------------------------------- F(2, 1488) = 14.16

Model | 10764.7166 2 5382.35829 Prob > F = 0.0000

Residual | 565719.09 1,488 380.187561 R-squared = 0.0187

-------------+---------------------------------- Adj R-squared = 0.0174

Total | 576483.807 1,490 386.901884 Root MSE = 19.498

-------------------------------------------------------------------------------

maxweeks | Coef. Std. Err. t P>|t| [95% Conf. Interval]

--------------+----------------------------------------------------------------

c\_age | .0343724 .0401705 0.86 0.392 -.0444245 .1131693

closebusiness | 1.192862 .2292293 5.20 0.000 .7432152 1.642509

\_cons | 5.650932 2.380279 2.37 0.018 .9818741 10.31999

-------------------------------------------------------------------------------

1. “Social distancing during the COVID-19 pandemic: Who are the present and future non-compliers?” <https://doi.org/10.7910/DVN/CMC00M> [↑](#footnote-ref-1)