Welcome!

Week 7: Regression

TA: Jennifer Lin

PS 210: Introduction to Empirical Methods

2021-11-03

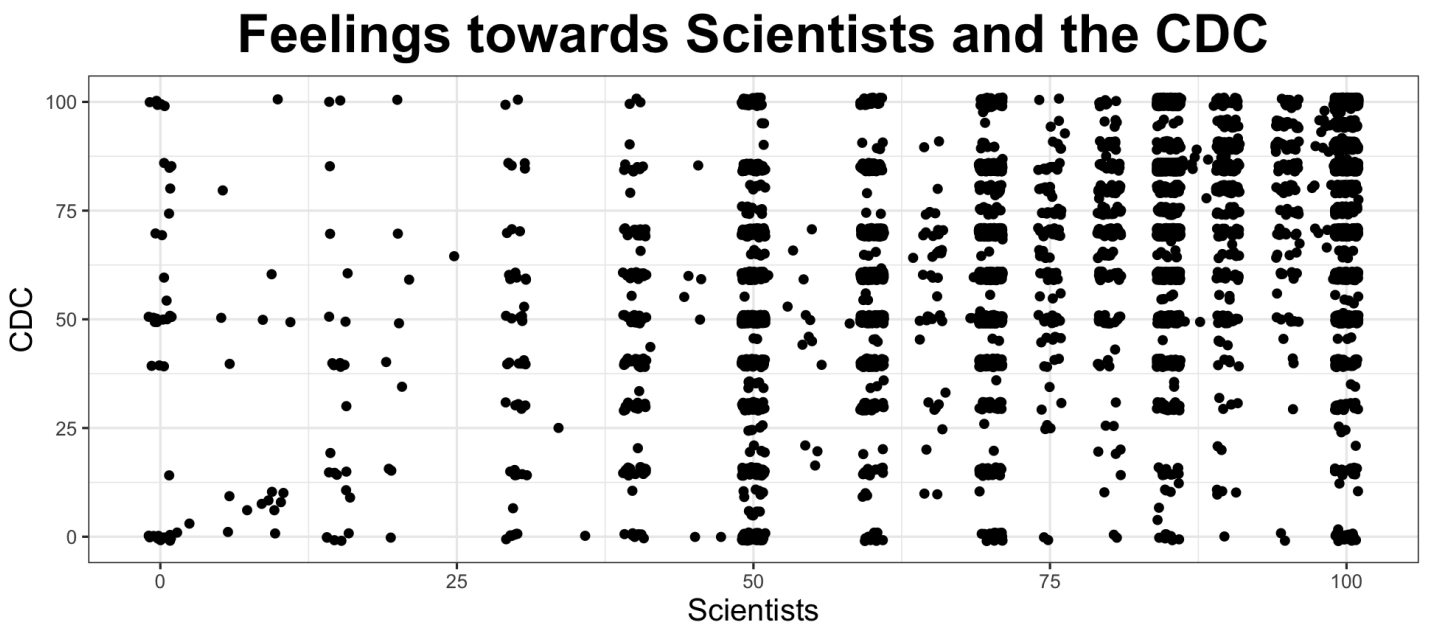
American National Elections Studies

Conducted every 4 years and measures a variety of political positions on various topics

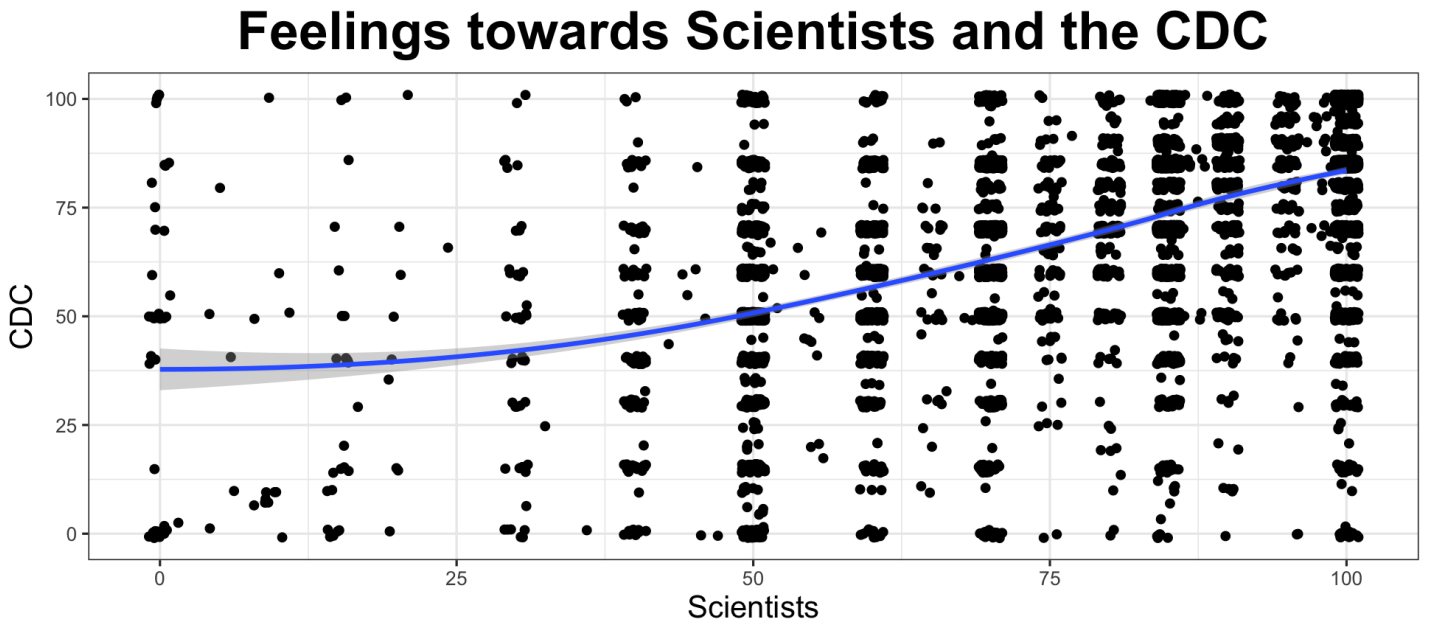
Specifically, they include a series of feeling thermometers that measure how warmly people feel towards a particular subject from a scale of 0 to 100.

Today, we will look at the feeling thermometers towards Scientists and the CDC.

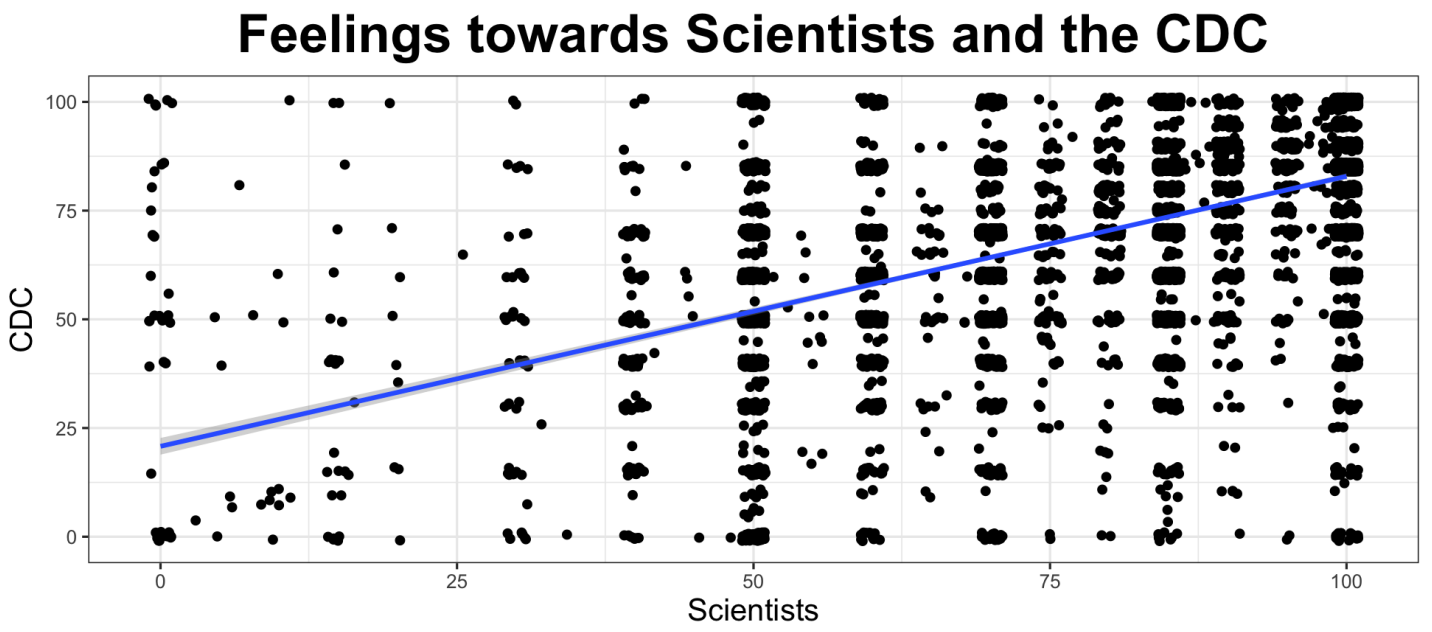
Scatter Plots



Fitting a Regression Line



Fitting a Regression Line



Correlation Between the variable

```
cor(ANES$FT_Scientist, ANES$FT_CDC, use = "complete.obs")
```

```
## [1] 0.5235262
```

Regression

```
##
## Call:
## lm(formula = FT_CDC ~ FT_Scientist, data = ANES)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -82.896 -11.846   2.104  14.322  79.205
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20.79532    0.97504   21.33   <2e-16 ***
## FT_Scientist  0.62101    0.01189   52.25   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.11 on 7230 degrees of freedom
##   (1048 observations deleted due to missingness)
## Multiple R-squared:  0.2741,    Adjusted R-squared:  0.274
## F-statistic:  2730 on 1 and 7230 DF,  p-value: < 2.2e-16
```

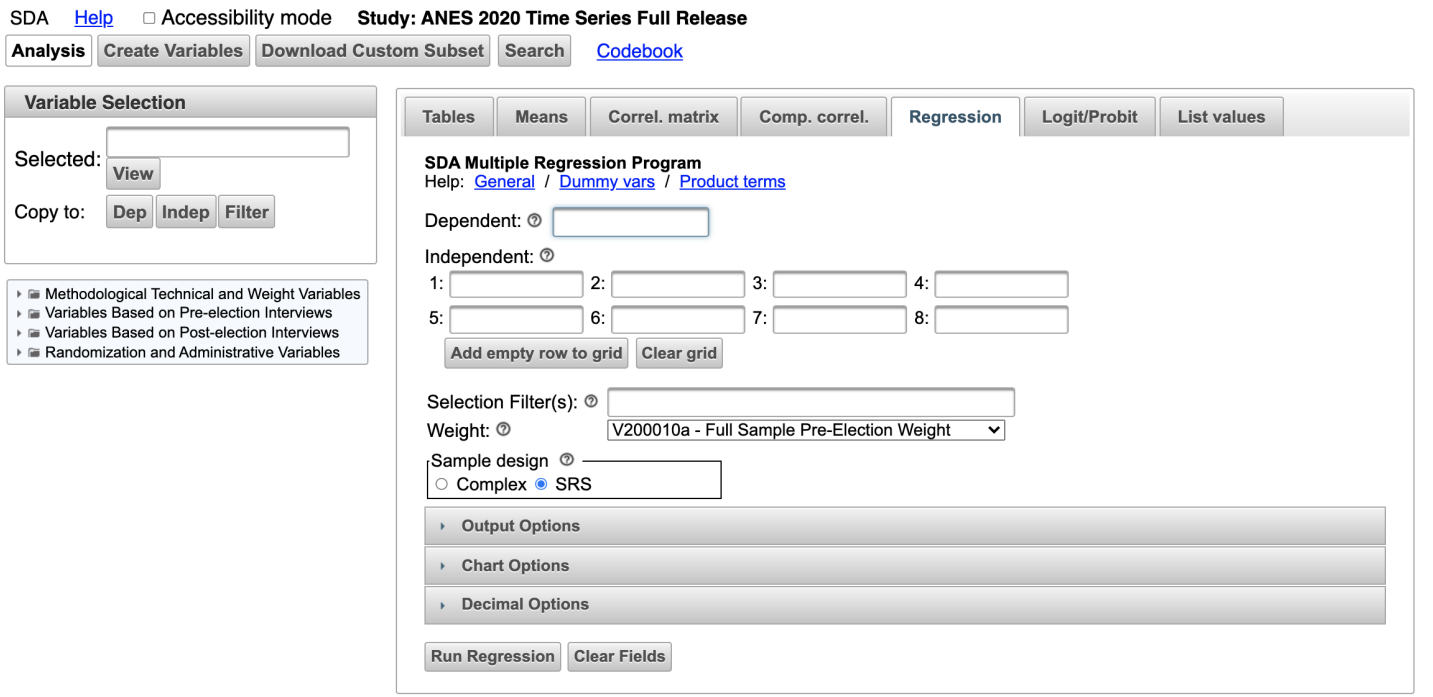
Interpretation

ON AVERAGE, for every one unit increase in feelings towards scientists, feelings towards the CDC increases by 0.62101

Why "ON AVERAGE"?

We are making a prediction from a line drawn to reflect the best fit (average) of the data.

Using the Software



SDA Help ☐ Accessibility mode **Study: ANES 2020 Time Series Full Release**

Analysis | Create Variables | Download Custom Subset | Search | Codebook

Variable Selection

Selected: [ ] View

Copy to: Dep | Indep | Filter

- Methodological Technical and Weight Variables
- Variables Based on Pre-election Interviews
- Variables Based on Post-election Interviews
- Randomization and Administrative Variables

Tables | Means | Correl. matrix | Comp. correl. | **Regression** | Logit/Probit | List values

SDA Multiple Regression Program
Help: General / Dummy vars / Product terms

Dependent: V202187

Independent:

1: V202173 2: 3: 4:

5: 6: 7: 8:

Add empty row to grid | Clear grid

Selection Filter(s):

Weight: V200010a - Full Sample Pre-Election Weight

Sample design
○ Complex ◉ SRS

- Output Options
- Chart Options
- Decimal Options

Run Regression | Clear Fields

Look at the Codebook

V202173	POST: FEELING THERMOMETER: SCIENTISTS
Question	How would you rate: Scientists
Value Labels	-9. Refused -7. No post-election data, deleted due to incomplete interview -6. No post-election interview -5. Interview breakoff (sufficient partial IW) -4. Technical error 998. Don't know
Survey Question(s)	THERMGR_SCIENT
Randomization	Set 1: Randomize the order of THGRFUND, THGRFEM, THGRLIB, THGRLAB, THGRBIGB, THGRCONS, THGRSCT, THGRGAY, THGRCONG, THGRMUSL, THGRXTIAN, JEWS, POLICE, TRANS, SCIENT, BLM, JOURN
Interviewer Instruction	{PROBE FOR DON'T KNOW RESPONSE: when you say don't know, do you mean that you don't know who this is or do you have something else in mind? ENTER number 0-100 }

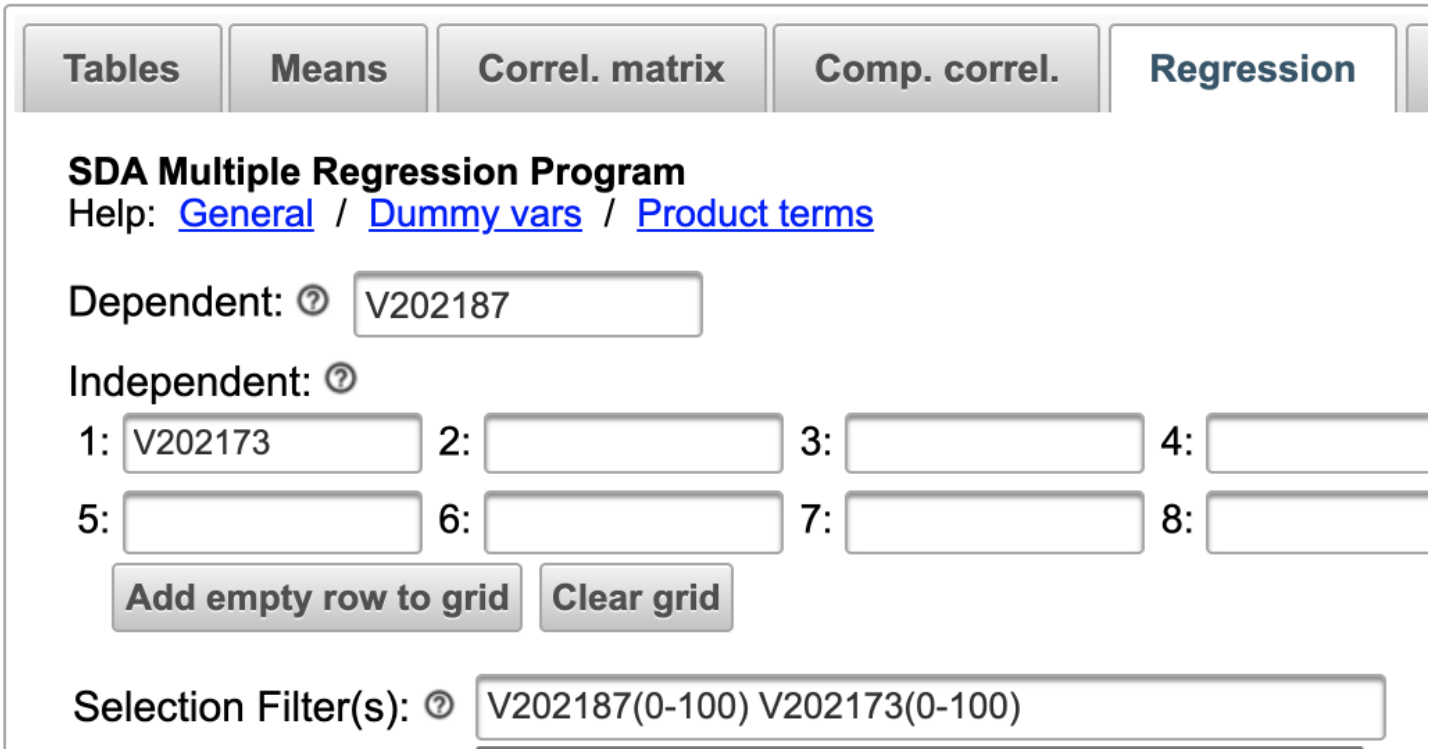
V202187	POST: FEELING THERMOMETER: CENTER FOR DISEASE CONTROL (CDC)
Question	How would you rate: The Centers for Disease Control (CDC)
Value Labels	-9. Refused -7. No post-election data, deleted due to incomplete interview -6. No post-election interview -5. Interview breakoff (sufficient partial IW) -4. Technical error 998. Don't know 999. Don't recognize
Survey Question(s)	THERMGR_CDC
Randomization	Set 2: Randomize the order of NATO, UN, NRA, SOCIAL, CAPITAL, FBI, ICE, METOO, RURAL, PLANPARENT, WHO, CDC
Interviewer Instruction	{PROBE FOR DON'T KNOW RESPONSE: when you say don't know, do you mean that you don't know who this is or do you have something else in mind? ENTER number 0-100 }

Notice...

There are some MISSING data that you might want to deal with.

The question accepts answers ranging from 0 - 100, with anything else being missing.

In the software, you should filter IN the observations so that the data do not reflect missing data


Tables
Means
Correl. matrix
Comp. correl.
Regression
SDA Multiple Regression Program
Help: General / Dummy vars / Product terms
Dependent: V202187
Independent:
1: V202173
2:
3:
4:
5:
6:
7:
8:
Add empty row to grid
Clear grid
Selection Filter(s): V202187(0-100) V202173(0-100)

SDA 4.1.3: Regression

ANES 2020 Time Series Full Release

Nov 03, 2021 (Wed 02:50 PM PDT)

Variables

Role	Name	Label	Range	MD	Dataset
Dependent	**V202187**	POST: Feeling thermometer: Center for Disease Control (CDC)	0-999		1
Independent	**V202173**	POST: Feeling thermometer: scientists	0-998		1
Filter	**V202187(0-100)**	POST: Feeling thermometer: Center for Disease Control (CDC)	0-999		1
Filter	**V202173(0-100)**	POST: Feeling thermometer: scientists	0-998		1

Regression Coefficients					Test That Each Coefficient = 0	
	B	SE(B)	Beta	SE(Beta)	T-statistic	Probability
V202173	.621	.012	.524	.010	52.247	.000
Constant	20.795	.975			21.328	.000

Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	T
Effect of each variable:	Negative			Positive			

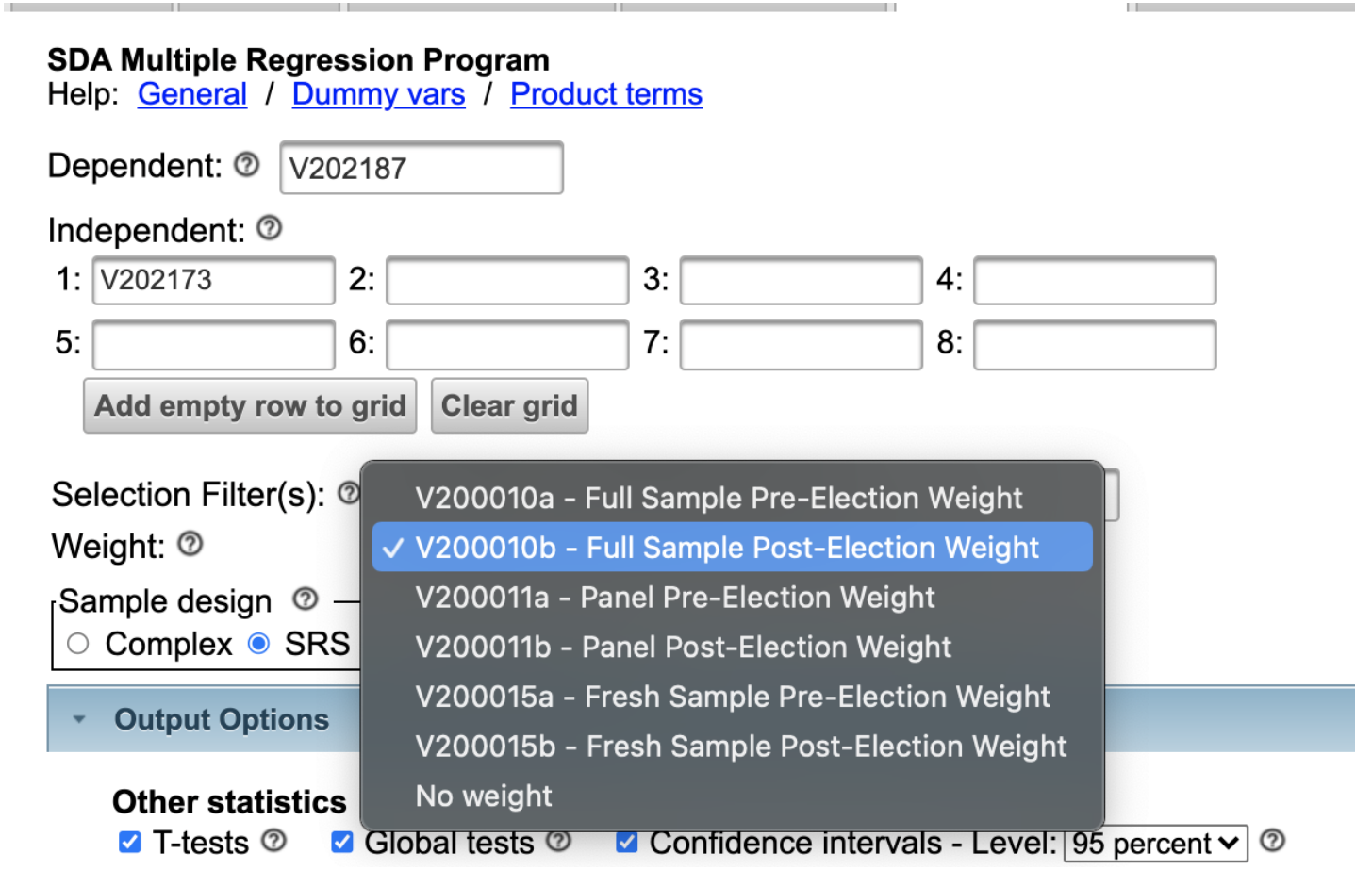
Multiple R = .524 R-Squared = .274 Adjusted R-Squared = .274 SE of Estimate (Root MSE) = 20.115

Global Tests for Groups of Variables

Group	Wald Chi-sq	df		Adjusted Wald F	P
		Numerator	Denominator		
All independent variables	2,729.771	1	7231	2,729.771	.000

Notice how it matches my results from before!

Adding Weights



SDA 4.1.3: Regression

ANES 2020 Time Series Full Release

Nov 03, 2021 (Wed 03:08 PM PDT)

Variables			
Role	**Name**	**Label**	
Dependent	**V202187**	POST: Feeling thermometer: Center for Disease Control (CDC)	
Independent	**V202173**	POST: Feeling thermometer: scientists	
Weight	**V200010b**	Full sample post-election weight	
Filter	**V202187(0-100)**	POST: Feeling thermometer: Center for Disease Control (CDC)	
Filter	**V202173(0-100)**	POST: Feeling thermometer: scientists	

Regression Coefficients					Test That Each Coefficient = 0	
	B	**SE(B)**	**Beta**	**SE(Beta)**	**T-statistic**	**Probability**
V202173	.574	.012	.491	.010	47.872	.000
Constant	24.453	.972			25.150	.000

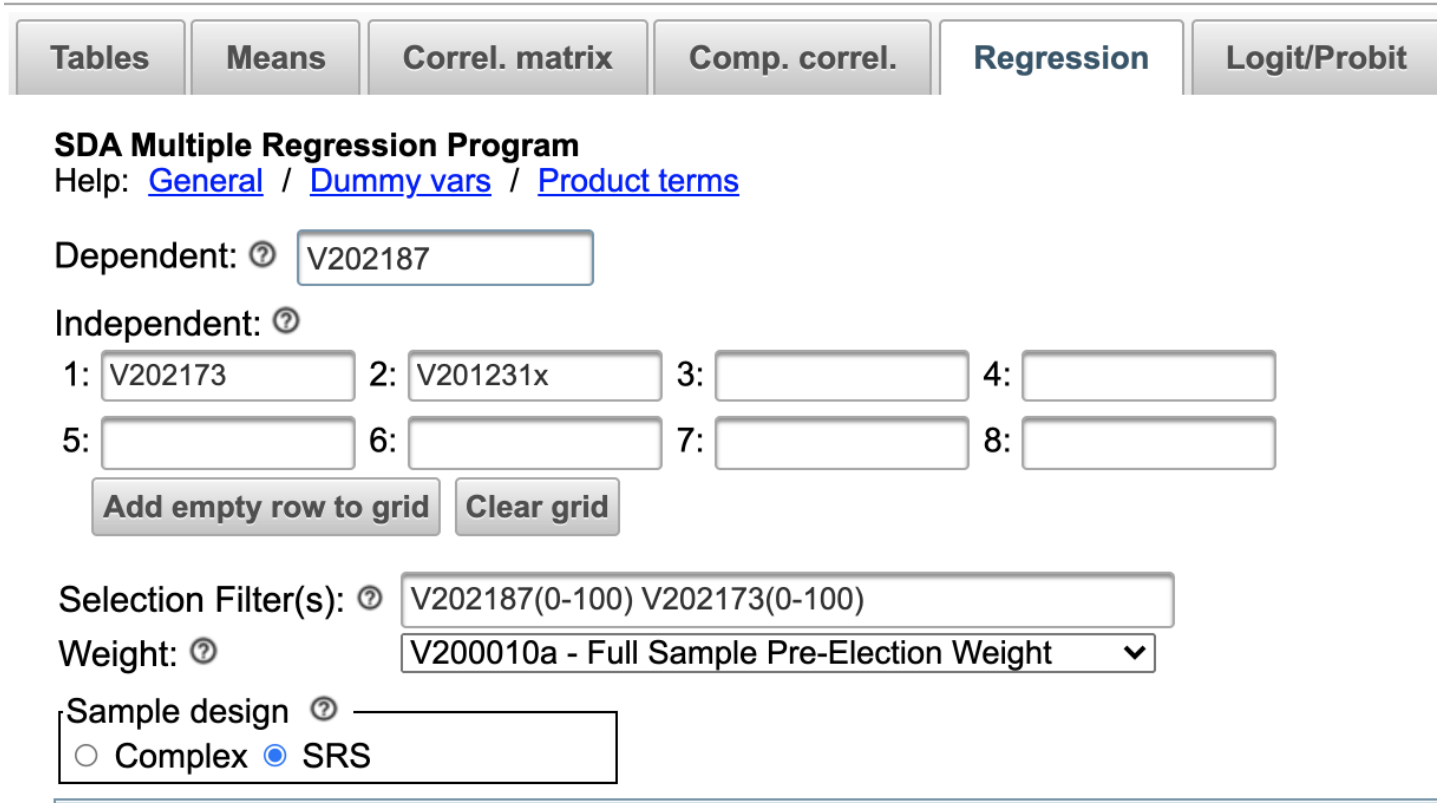
Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	T
Effect of each variable:	Negative			Positive			

Multiple R = .491 R-Squared = .241 Adjusted R-Squared = .241 SE of Estimate (Root MSE

Notice the effects of adding weights

The estimates change ever so slightly as you are accounting for biases in the sampling process

Adding Confounders -- Like Partisanship



SDA 4.1.3: Regression

ANES 2020 Time Series Full Release

Nov 03, 2021 (Wed 03:00 PM PDT)

Variables		
Role	**Name**	**Label**
Dependent	**V202187**	POST: Feeling thermometer: Center for Disease Control (CDC)
Independent	**V202173**	POST: Feeling thermometer: scientists
Independent	**V201231x**	PRE: SUMMARY: Party ID
Weight	**V200010a**	Full sample pre-election weight
Filter	**V202187(0-100)**	POST: Feeling thermometer: Center for Disease Control (CDC)
Filter	**V202173(0-100)**	POST: Feeling thermometer: scientists

Regression Coefficients					Test That Each Coefficient = 0	
	B	**SE(B)**	**Beta**	**SE(Beta)**	**T-statistic**	**Probability**
V202173	.514	.013	.435	.011	40.543	.000
V201231x	-1.979	.114	-.185	.011	-17.296	.000
Constant	36.929	1.259			29.341	.000

Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	T
Effect of each variable:	Negative			Positive			

Template for Regression Estimates Interpretations

On average, for every one unit increase in [X], [Y] [increases/decreases] by [ESTIMATE]

This is intuitive, if you think about the algebra of a linear model

$$[Y] = [ESTIMATE] * [X] + error$$

If [X] is 1, the [Y] increases by the value of the [ESTIMATE]