POLS2044 WEEK 10
Multivariate regression

Australian National University
School of Politics & International Relations
Dr. Richard Frank

In Week 10 of POLS2044 we will be focusing on ways of testing hypotheses related to the relationships between two variables in a regression framework while controlling for additional theoretically important explanatory factors. This week builds directly on the last week's hypothesis tests about bivariate regressions.

This week I have three main goals. First, I want to (1) introduce you to multivariate regression analysis, (2) explain why it is necessary to pass the fourth hurdle, and (3) explain its assumptions. Second, I want to work through how we might interpret important regression results. Third, I want to give you an opportunity to practice generating and explaining your regression results.

Reading notes

There is one assigned reading, Chapter 10 from the textbook. As in the last few weeks, there are myriad formulae in the readings that may be a bit difficult for some. Remember, I am interested in you being able to understand why we are using different metrics, what elements go into these metrics, how these metrics change as the values of the inputs change, and what their results allow us to conclude about what we care about (answering our research questions).

LECTURE PART 1: Multivariate regression

Today's motivating questions

Why do we need to move from bivariate to multivariate regression?
How do we do so?
How do we interpret our results?

The four causal hurdles

1. A credible causal mechanism
2. Ruling out endogeneity
3. Covariation
4. Controlling for confounding variables that may make current association spurious?

Source: Kelstedt & Whitten (2018: 55)

Moving from bivariate to multivariate regression

Taking this step to multiple regression finally enables us to potentially pass the fourth hurdle for the first* time.

* not counting experimental methods

Controlling for other factors

Experimental research designs control for other theoretically relevant factors through random assignment into treatment and control groups.

This is the gold standard.

Observational research designs control for other factors by adding them to the regression model.

This involves (1) theoretically identifying relevant factors (e.g., culture) and (2) finding observable/measurable indicators of them.

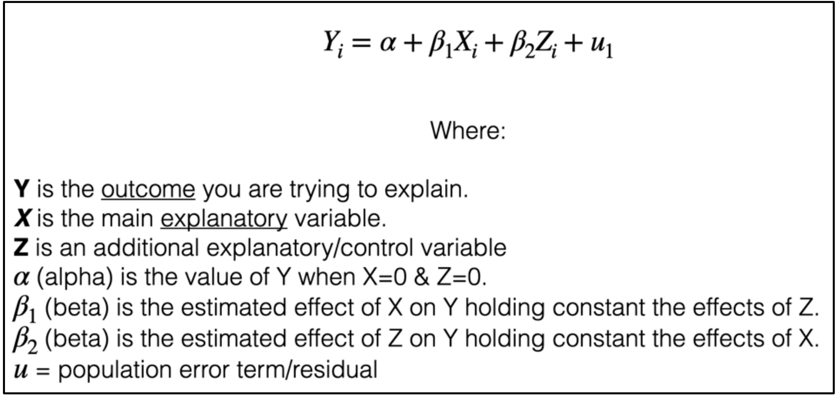
Bivariate regression model

$$\boxed{Y_i = \alpha + \beta X_i + u_i}$$

Where:

Y is the outcome you are trying to explain.
X is the main explanatory variable.
(alpha) is the value of Y when X=0.
(beta) is the estimated relationship between X and Y.
u is the population error term/residual

Multivariate regression model



$$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_1$$

Where:

Y is the outcome you are trying to explain.
X is the main explanatory variable.
Z is an additional explanatory/control variable
α (alpha) is the value of Y when X=0 & Z=0.
β_1 (beta) is the estimated effect of X on Y holding constant the effects of Z.
β_2 (beta) is the estimated effect of Z on Y holding constant the effects of X.
u = population error term/residual

Three things worth noting

1. Subscripts

Bivariate——$Y_i = \alpha + \beta X_i + u_i$
Multivariate—$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_i$

Notice the subscripts for the variables and the slope coefficients (￼s).
The subscript “i” tells us that the equation is for each observation of our variables from i to n.

Variables by themselves (e.g., Y, X, Z) actually represent a vector (i.e.values of each variable).

2. Estimated Beta1 is affected by inclusion of Z

Bivariate——$Y_i = \alpha + \beta X_i + u_i$
Multivariate—$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_1$

Adding in $\beta_2 Z_i$ to the equation changes the estimation of $\beta_1 X_i$, sometimes by a little, sometimes a lot.

Four possible changes

1. β_1 was statistically significant but is **no longer statistically significant**
2. β_1 was not statistically significant but is **now statistically significant**
3. β_1 value is larger than before (i.e., potentially **more substantively significant**)
4. β_1 value is smaller than before (i.e., potentially **less substantively significant**)

Two-variable regression slope: $\beta = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}(X_i - \bar{X})^2}$

Multivariate regression slope: $\widehat{\beta_1} = \dfrac{\sum_{i=1}^{n}(X_i - \widehat{X_i})(Y_i - \widehat{Y_i^*})}{\sum_{i=1}(X_i - \widehat{X_i})^2}$

where $\widehat{Y}_i^* = \hat{\alpha}^* + \hat{\beta}^* Z_i$

And $\widehat{X_i}$ is the predicted value of X based on Z.

3. Avoid multicollinearity

Perfect multicollinearity is when there is an linear relationship between two independent variables.

Some examples of perfect multicollinearity include spatial (e.g., EU country/non-EU country) or temporal (Cold War/Post-Cold War) dummy variables.

If there is perfect multicollinearity, one of the variables will be automatically dropped by your software.

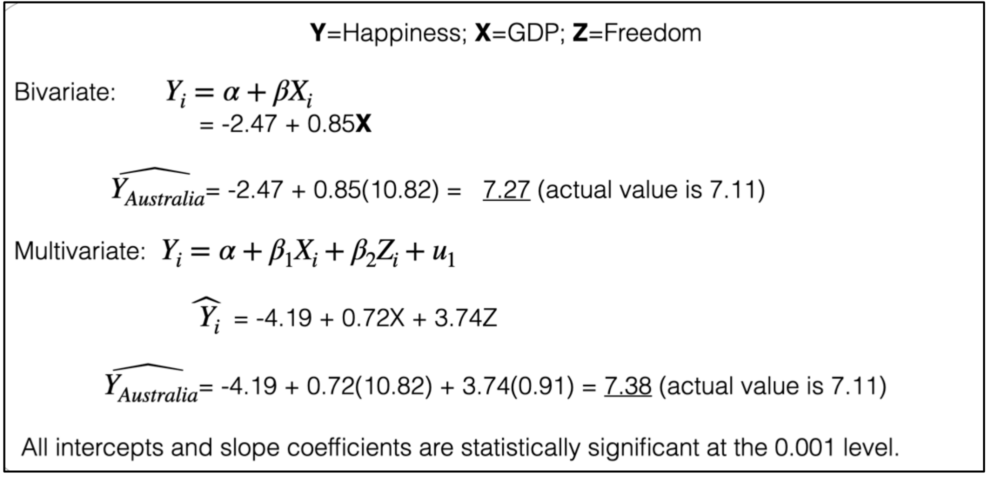
If there is high collinearity between independent variables, this will distort your parameter estimates.

There are tests for multicollinearity (e.g., variance inflation factors), but initially I would suggest creating a correlation table of your independent variables.

LECTURE PART 2: Interpreting multiple regression results

Now we have some regression results, what do we do with them?
Let’s continue with the happiness data example from last week.

Estimating the relationship between X and Y, controlling for Z



Y=Happiness; **X**=GDP; **Z**=Freedom

Bivariate: $Y_i = \alpha + \beta X_i$

$= -2.47 + 0.85\mathbf{X}$

$\widehat{Y_{Australia}} = -2.47 + 0.85(10.82) =$ $\underline{7.27}$ (actual value is 7.11)

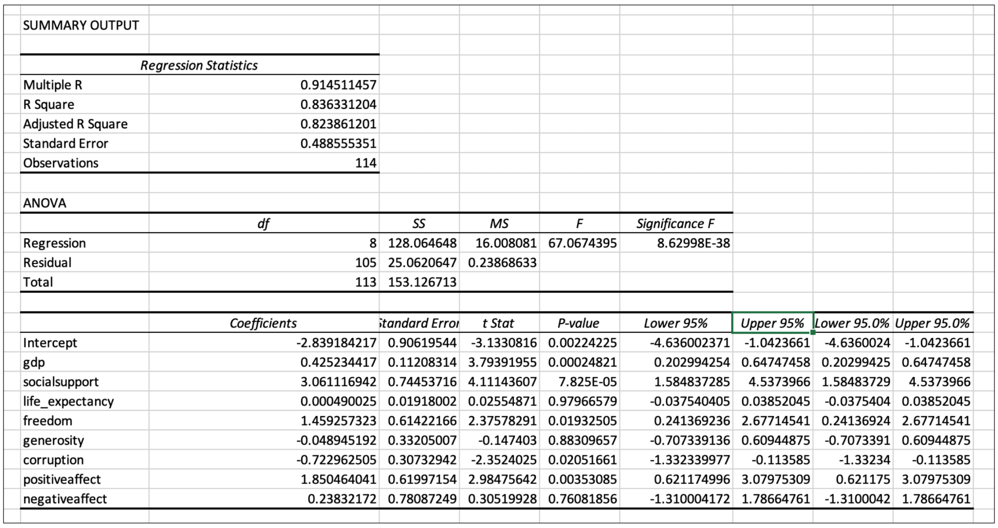
Multivariate: $Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_1$

$\widehat{Y_i} = -4.19 + 0.72X + 3.74Z$

$\widehat{Y_{Australia}} = -4.19 + 0.72(10.82) + 3.74(0.91) =$ $\underline{7.38}$ (actual value is 7.11)

All intercepts and slope coefficients are statistically significant at the 0.001 level.

More generally, how do we interpret this regression table?



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.914511457
R Square	0.836331204
Adjusted R Square	0.823861201
Standard Error	0.488555351
Observations	114

ANOVA

	df	SS	MS	F	Significance F
Regression	8	128.064648	16.008081	67.0674395	8.62998E-38
Residual	105	25.0620647	0.23868633		
Total	113	153.126713			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-2.839184217	0.90619544	-3.1330816	0.00224225	-4.636002371	-1.0423661	-4.6360024	-1.0423661
gdp	0.425234417	0.11208314	3.79391955	0.00024821	0.202994254	0.64747458	0.20299425	0.64747458
socialsupport	3.061116942	0.74453716	4.11143607	7.825E-05	1.584837285	4.5373966	1.58483729	4.5373966
life_expectancy	0.000490025	0.01918002	0.02554871	0.97966579	-0.037540405	0.03852045	-0.0375404	0.03852045
freedom	1.459257323	0.61422166	2.37578291	0.01932505	0.241369236	2.67714541	0.24136924	2.67714541
generosity	-0.048945192	0.33205007	-0.147403	0.88309657	-0.707339136	0.60944875	-0.7073391	0.60944875
corruption	-0.722962505	0.30732942	-2.3524025	0.02051661	-1.332339977	-0.113585	-1.33234	-0.113585
positiveaffect	1.850464041	0.61997154	2.98475642	0.00353085	0.621174996	3.07975309	0.621175	3.07975309
negativeaffect	0.23832172	0.78087249	0.30519928	0.76081856	-1.310004172	1.78664761	-1.3100042	1.78664761

How do we interpret this table?

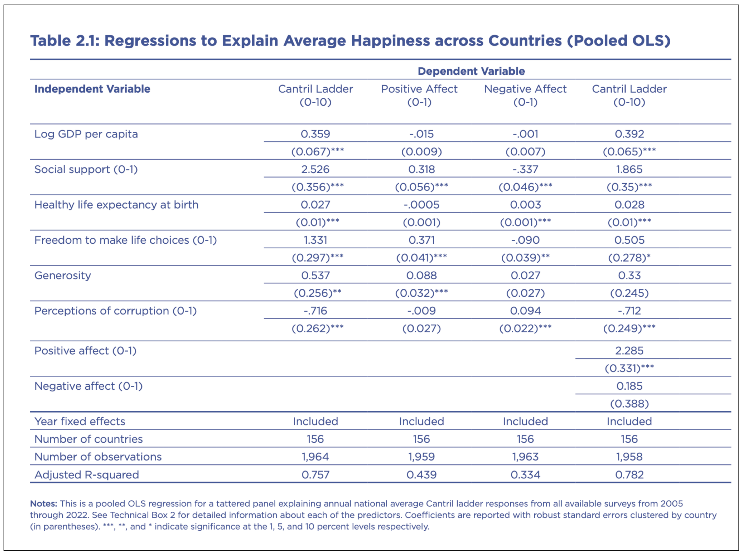


Table 2.1: Regressions to Explain Average Happiness across Countries (Pooled OLS)

	Dependent Variable			
Independent Variable	Cantril Ladder (0-10)	Positive Affect (0-1)	Negative Affect (0-1)	Cantril Ladder (0-10)
Log GDP per capita	0.359	-.015	-.001	0.392
	(0.067)***	(0.009)	(0.007)	(0.065)***
Social support (0-1)	2.526	0.318	-.337	1.865
	(0.356)***	(0.056)***	(0.046)***	(0.35)***
Healthy life expectancy at birth	0.027	-.0005	0.003	0.028
	(0.01)***	(0.001)	(0.001)***	(0.01)***
Freedom to make life choices (0-1)	1.331	0.371	-.090	0.505
	(0.297)***	(0.041)***	(0.039)**	(0.278)*
Generosity	0.537	0.088	0.027	0.33
	(0.256)**	(0.032)***	(0.027)	(0.245)
Perceptions of corruption (0-1)	-.716	-.009	0.094	-.712
	(0.262)***	(0.027)	(0.022)***	(0.249)***
Positive affect (0-1)				2.285
				(0.331)***
Negative affect (0-1)				0.185
				(0.388)
Year fixed effects	Included	Included	Included	Included
Number of countries	156	156	156	156
Number of observations	1,964	1,959	1,963	1,958
Adjusted R-squared	0.757	0.439	0.334	0.782

Notes: This is a pooled OLS regression for a tattered panel explaining annual national average Cantril ladder responses from all available surveys from 2005 through 2022. See Technical Box 2 for detailed information about each of the predictors. Coefficients are reported with robust standard errors clustered by country (in parentheses). ***, **, and * indicate significance at the 1, 5, and 10 percent levels respectively.

Source: World Happiness Report 2023: 38

How do we compare results tables?

LECTURE PART 3: How do we create a regression table?

First, theoretically…

Theoretically, you want to:
(1) build on the best existing research and show you can replicate/approximate it,
(2) demonstrate whether your results support or fail to support your alternate hypothesis(es), and
(3) demonstrate whether or not your results are robust to alternate theoretical and practical specifications.

Then practically.

You want to include enough information to allow readers to:
(1) understand what you did,
(2) reach their own conclusions as to whether your results are statistically and substantively significant,
(3) replicate your research if they are interested.

Replication is crucial to scientific progress.

Example from Reinhart and Rogoff

Source: https://theconversation.com/the-reinhart-rogoff-error-or-how-not-to-excel-at-economics-13646

2023 winner of Nobel Prize for Medicine: Katalin Karikó

Demoted from U.Penn in 1995 when unable to secure grants. “Not of faculty quality”.
Seminal paper desk rejected from *Nature* in 2005.

1. Choose your dependent variables

Research is an often messy, time-intensive, stressful, and confusing process.
There are usually multiple ways to define your dependent variable (e.g., continuous, dichotomous, change, logged).
Often people will study multiple variations of their dependent variable and only report one’s results.

2. Choose independent variables & control variables

What is/are your main independent variables?
What are other factors (i.e., control variables) that theoretically affect your outcome?
What are the most theoretically grounded way of measuring these factors (e.g., absolute value, % GDP, % population, logged)?

3. **Decide what models are important to summarise**

   Usually, there is a standard/influential model in your research area.
   Report your replication of those results.
   Then compare the results with your best model.
   Add additional models to incorporate other hypotheses, control variables, or methodological concerns.

4. **Check your sample**

   Make sure your sample includes what you think it includes.
   Think about how it represents/fails to represent the population you are theorising about.
   Explore the data for potential outliers or cases with missing data.
   Think about important cross-temporal or cross-spatial differences.

5. **Then turn something like the Excel results into a table**

6. **Interpret the table**

   Tell your readers in words what you want them to take away from your table.
   Often focus is on both statistical and substantive significance.
   Connect results back to your theory and hypotheses.

Today's motivating questions

Why do we need to move from bivariate to multivariate regression?
How do we do so?
How do we interpret our results?

Important Week 10 terms

Bias
Omitted variable bias
Perfect multicollinearity
Substantive significance
Vector
Matrix

WEEK 10 WORKSHOP

Welcome to our penultimate workshop! Today we are going to be focusing on running and interpreting multiple regressions. **Remember to submit your own work (not anyone else's) to "Wattle/Week 10/Workshop/Item 9.1" at the end of workshop**.

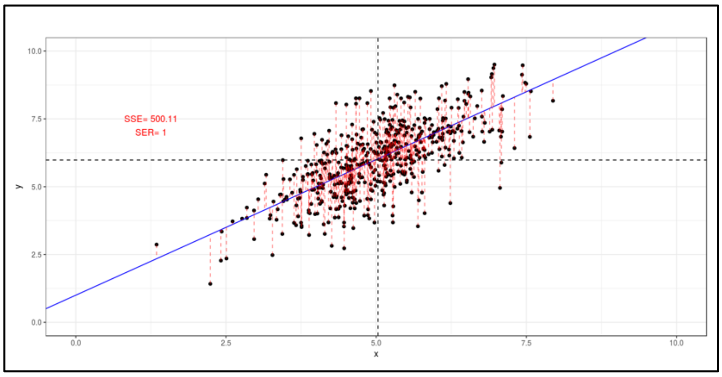
Part 1: Visualising regression lines

This website has clear capacity limits. Please complete this section if you can get access. Skip to Part 2 if you cannot.

Today we are going to start off by reinforcing our understanding of what an ordinary least squares regression is doing under the hood. More specifically, we are going to see how choosing different regression lines through our observations changes the sum of squared errors (SSE). Remember, the whole (technical) point of regression is to try and minimize the sum of squared errors.

***Step 1*:** Go to the following website:
https://ryansafner.shinyapps.io/ols_estimation_by_min_sse/

What you see is a graph plotting a series of observations according to their X and Y values.



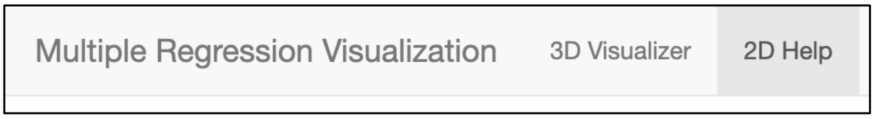
On the right are two sliders for the slope (labelled "a" here but "beta" in the readings and lecture) and for the intercept ("b" here but "alpha" in the readings and lecture). As you drag the sliders left and right you will see how the blue regression line shifts and the red dotted lines that represent the residuals. Normally we want to try and minimise the SSE, but today I also want to see what the highest SSE you can find.

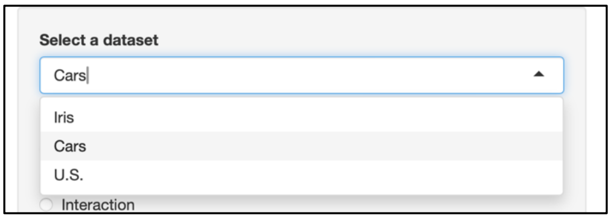
Step 2: Adjust the sliders for slope and intercept to try and maximise and minimise the SSE.

> **Question #1: What are the values of the slope and intercept that give you the smallest SSE you can find?**
>
> **Question #2: What are the values of the slope and intercept that give you the largest SSE you can find?**

Let's kick it up a notch. Let's try and run a regression with two independent variables.

***Step 3**:* Go to the following website: https://calpolystat2.shinyapps.io/3d_regression/. Click on the “2D Help” tab at the top. Then select the “Cars” dataset using the dropdown menu on the left.





What you should see now is a graph of a sample of American cars. Specifically, their weight (*wt*) and fuel economy (*mpg*). If you do not see the figure, you may need to (1) allow JavaScript to run on this website or (2) click your mouse/trackpad on the empty space below “Plot.”

Hopefully, you see a negatively sloped line. As weight increases, the expected fuel economy declines. Look at the “Model Info” tab and you will see the regression results of fuel economy (“miles per gallon” regressed on weight (“wt”).

This is a pretty minimal model of fuel economy. We can do better.

Step 4: Click on “3D visualiser” at the top of the page. Select “Cars” again in the dataset selection dropdown menu, then select “Horsepower + Weight” under “Available Models”. Click on the “Model Info” tab.

> **Question #3: Was there any change in wt’s t-value and p-value when we move from a one independent variable model (X=*wt*) to the two independent model (X=*wt* and Z=*hp*)? Was there a change in the *wt* slope coefficient? In what direction? What about the estimated R^2?**

Step 5: Now look at the Plot tab. Drag the cube around until you see mpg on the horizontal axis and weight on the vertical axis. Then drag the cube until horsepower is on the horizontal axis and mpg is on the vertical axis.

Notice how the estimated regression line is now represented by a flat plane. As you rotate the cube you can see both the observations and the plane rotate around.

> **Question #4: Does the sign of the two slope coefficients match the direction of the plane when you rotate the cube so that either your X or Y are on the horizontal axis? If so, how?**

If you really want to see something cool, try clicking on the available model “Horsepower + Weight + Transmission.

Question #5: Given how the cube changes when you switch to three independent variables, what can you tell about what values "transmission" are likely to take on?

This is the end of Part 1. Hopefully, this exercise helped you visualise the relationship between our variables a bit differently and connect the model results to the data distribution. This part focused on trying to strengthen our grasp of the intuition behind how regression works. Now we are going to turn in Part 2 to developing a (theoretical and empirical) multivariate model before we test a few models in Part 3.

Part 2: Developing a multivariate model

In lecture, I stressed the importance of thinking theoretically about (1) how our research design ties back to our research question and theories, (2) how we can build on the best existing research, and (3) how we can ensure that our results are robust to alternate theoretical and practical model specifications. I also ran through the different steps in running a regression and turning the results into a table.

Last week we all used the same dependent variable (`happiness`) but analysed the effects of different independent variables. This week, we will be starting by using the same group of dependent variables from a new dataset. We will also be using new data centring on the 2022 Australian election from the Australian Election Study. You can read more about the survey at https://australianelectionstudy.org/.[1] The original dataset had dozens of variables and thousands of observations. I cleaned up the data by renaming the variables I wanted to focus on, removing respondents who did not answer every question, and keeping every ninth observation, which left me with 208 observations. **The data are available on Wattle under Week 10.**

Step 6: Choose your dependent variable by rolling a die.

Question #6: what is your dependent variable?

The number you roll is the number you choose from the table below. Please make sure that everyone at your table chooses a different number. If you roll the same number as another student roll again until you get a number not previously rolled by another student.

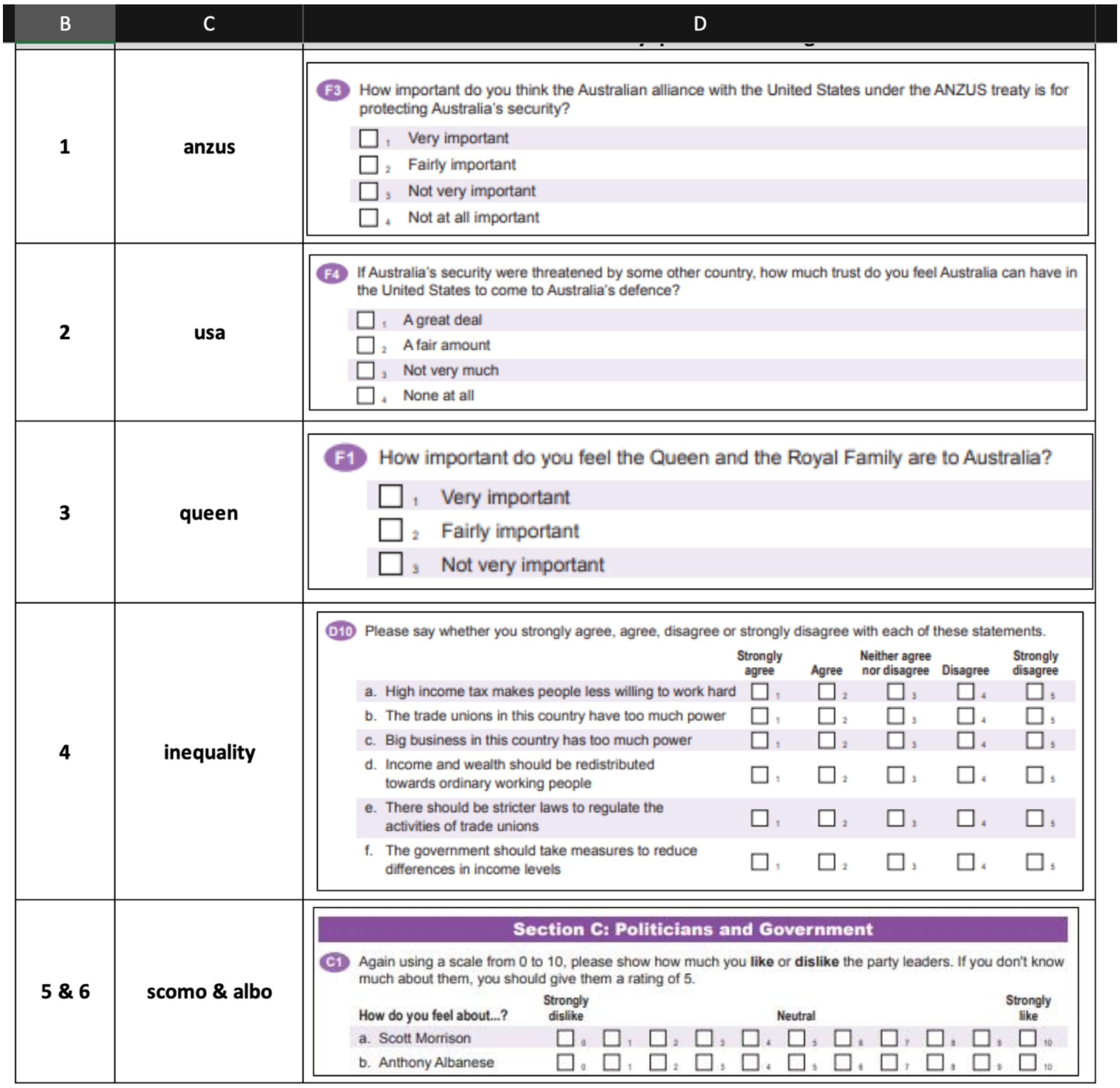
Dependent variables

Die roll	Dependent variable
1	*anzus*
2	*usa*
3	*queen*
4	*inequality*
5	*scomo*
6	*albo*

Each of these variables are described in the table below, which is also in the spreadsheet. Read the question wording and think a bit about what might affect people's answers to your question.

[1] The data were downloaded from https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.26193/W3U2S3 .

Look at the numbers by the checked boxes. These are the numbers coded in the spreadsheet you will be analysing.



B	C	D
1	anzus	F3 How important do you think the Australian alliance with the United States under the ANZUS treaty is for protecting Australia's security? □ 1 Very important □ 2 Fairly important □ 3 Not very important □ 4 Not at all important
2	usa	F4 If Australia's security were threatened by some other country, how much trust do you feel Australia can have in the United States to come to Australia's defence? □ 1 A great deal □ 2 A fair amount □ 3 Not very much □ 4 None at all
3	queen	F1 How important do you feel the Queen and the Royal Family are to Australia? □ 1 Very important □ 2 Fairly important □ 3 Not very important
4	inequality	D10 Please say whether you strongly agree, agree, disagree or strongly disagree with each of these statements. Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree a. High income tax makes people less willing to work hard □ 1 □ 2 □ 3 □ 4 □ 5 b. The trade unions in this country have too much power □ 1 □ 2 □ 3 □ 4 □ 5 c. Big business in this country has too much power □ 1 □ 2 □ 3 □ 4 □ 5 d. Income and wealth should be redistributed towards ordinary working people □ 1 □ 2 □ 3 □ 4 □ 5 e. There should be stricter laws to regulate the activities of trade unions □ 1 □ 2 □ 3 □ 4 □ 5 f. The government should take measures to reduce differences in income levels □ 1 □ 2 □ 3 □ 4 □ 5
5 & 6	scomo & albo	Section C: Politicians and Government C1 Again using a scale from 0 to 10, please show how much you **like** or **dislike** the party leaders. If you don't know much about them, you should give them a rating of 5. How do you feel about...? Strongly dislike / Neutral / Strongly like a. Scott Morrison □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □ 8 □ 9 □ 10 b. Anthony Albanese □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □ 8 □ 9 □ 10

Step 7: Choose one primary independent variable and at least one control variable.

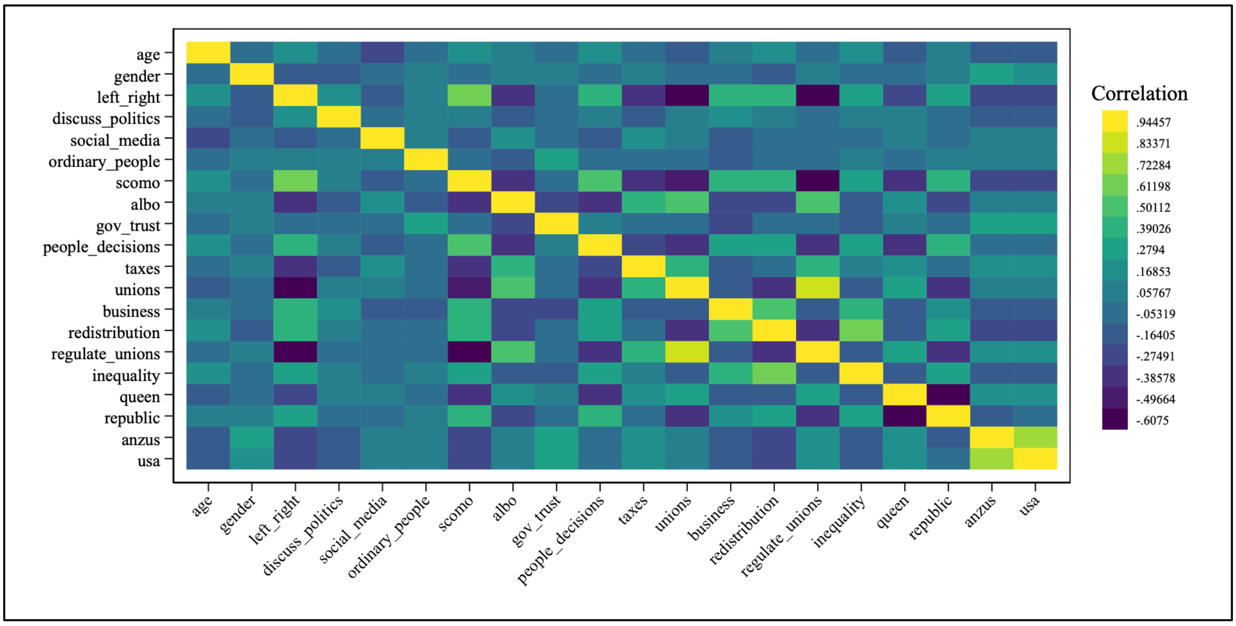
These variables are ones you will be using in your Excel regression model. Think carefully about your theoretical and empirical research design. Below are three safe choices for independent variables. In my tests at least one of these variables is a statistically significant predictor of all six outcome variables in the table above.

Safe choices for independent variables

Variable	**description**
age	Age of respondent
gender	Male=1; female=2; other=3
left_right	Left-right scale (0=left to 10=right)

However, there are myriad other interesting variables included in the data I have cleaned and uploaded. Their descriptions are in the "independent variables" tab in the spreadsheet. I have included a correlation matrix of all variables as well as a table of descriptive statistics. Looking at these may help you understand the data better as well as give you some ideas of what variables you want to explore.

Correlation matrix



Remember that correlation coefficients run from -1 to 1 with 0 meaning the variables do not covary at all, -1 means that they are perfectly negatively correlated (as one increases/decreases the other decreases/increases). The legend on the right of the figure shows the threshold correlation values between different colours.

Descriptive statistics for all variables in our dataset

Variable	Obs	Mean	Standard deviation	Minimum	Maximum
age	208	55.08	17.03	18	89
gender	208	1.53	0.51	1	3
left_right	208	4.43	2.18	0	10
discuss_politics	208	1.70	0.80	0	3
social_media	208	0.63	0.76	0	3
ordinary_people	208	3.19	1.11	1	5
scomo	208	2.99	3.27	1	10
albo	208	6.06	2.56	0	10
gov_trust	208	2.13	0.72	1	4
people_decisions	208	1.50	0.71	1	4
taxes	208	3.08	1.15	1	5
unions	208	2.99	1.15	1	5
business	208	1.95	0.87	1	5
redistribution	208	2.61	1.14	1	5
regulate_unions	208	2.88	1.10	1	5
inequality	208	2.48	1.07	1	5
queen	208	2.45	0.74	1	3
republic	208	2.26	1.01	1	4
anzus	208	1.67	0.79	1	4
usa	208	1.20	0.81	1	4

Question #7: What is your primary explanatory variable? What is a plausible causal mechanism linking this variable to your outcome variable?

Question #8: What is your null hypothesis and alternate hypothesis?

Question #9: What is/are your control variable(s)?

Part 3: Running and analysing a multivariate model

Now that you have the main elements of a regression, now is the time to run your model and analyse your results. *You may find it easier to run regressions if you cut and paste your independent variables so that they are in neighbouring columns.*

Step 8: Run a regression including only your main explanatory variable (X) and your outcome variable (Y). Paste/put the outcome into the "analysis" sheet.

Step 9: Run a regression including your main explanatory variable (X) AND your control variable (Z). Paste/put the outcome into the "analysis" sheet.

Question #10: Make a regression results table akin to what you have seen in the lectures and readings this term. Include two columns, one for each regression you run.

Make sure you include the following elements: slope coefficients (and their standard errors), intercepts (and their standard errors), indicators of statistical significance (e.g., stars), sample size, F-statistics (and p-value stars if they are significant), and R^2. Also make sure to only include two or three decimal places. If a value is very small (e.g., 5.7E-13), just put 0.00 or 0.000.

Here is an example from Kostadinova and Power (2007: 368).[2]

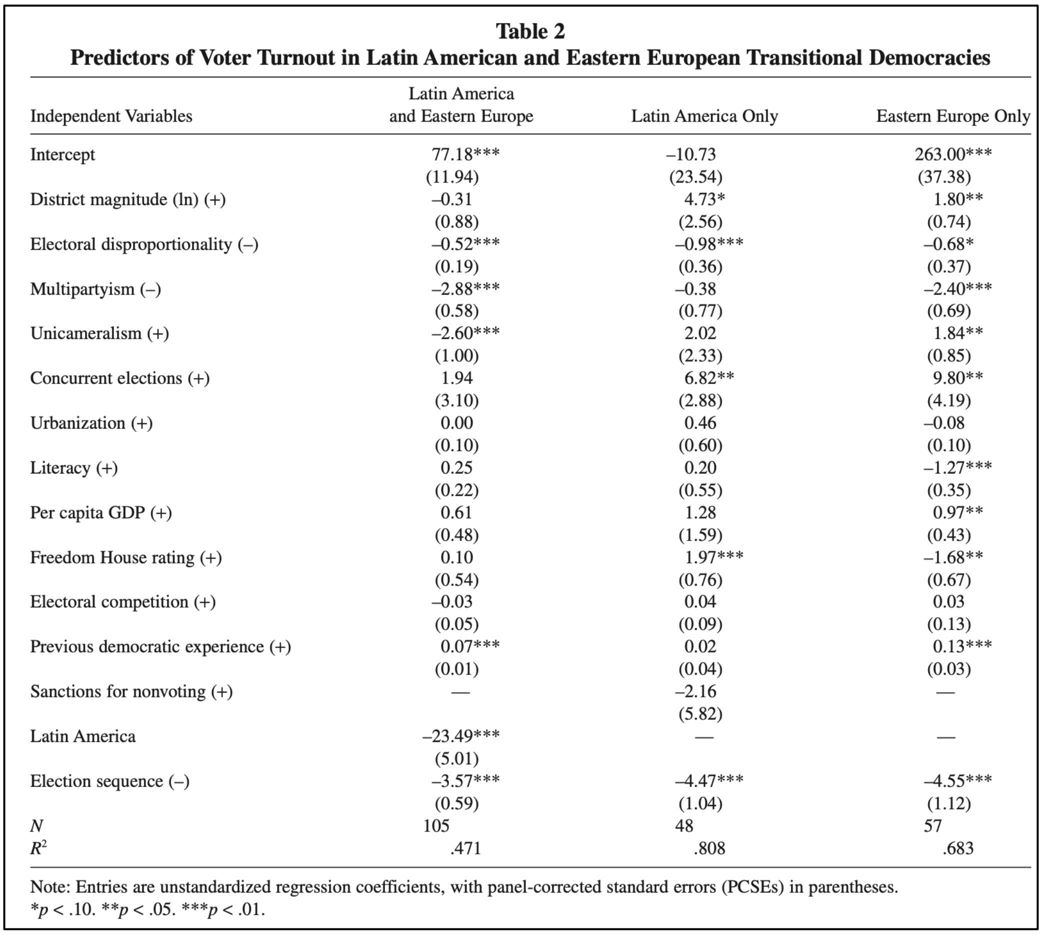


Table 2
Predictors of Voter Turnout in Latin American and Eastern European Transitional Democracies

Independent Variables	Latin America and Eastern Europe	Latin America Only	Eastern Europe Only
Intercept	77.18***	–10.73	263.00***
	(11.94)	(23.54)	(37.38)
District magnitude (ln) (+)	–0.31	4.73*	1.80**
	(0.88)	(2.56)	(0.74)
Electoral disproportionality (–)	–0.52***	–0.98***	–0.68*
	(0.19)	(0.36)	(0.37)
Multipartyism (–)	–2.88***	–0.38	–2.40***
	(0.58)	(0.77)	(0.69)
Unicameralism (+)	–2.60***	2.02	1.84**
	(1.00)	(2.33)	(0.85)
Concurrent elections (+)	1.94	6.82**	9.80**
	(3.10)	(2.88)	(4.19)
Urbanization (+)	0.00	0.46	–0.08
	(0.10)	(0.60)	(0.10)
Literacy (+)	0.25	0.20	–1.27***
	(0.22)	(0.55)	(0.35)
Per capita GDP (+)	0.61	1.28	0.97**
	(0.48)	(1.59)	(0.43)
Freedom House rating (+)	0.10	1.97***	–1.68**
	(0.54)	(0.76)	(0.67)
Electoral competition (+)	–0.03	0.04	0.03
	(0.05)	(0.09)	(0.13)
Previous democratic experience (+)	0.07***	0.02	0.13***
	(0.01)	(0.04)	(0.03)
Sanctions for nonvoting (+)	—	–2.16	—
		(5.82)	
Latin America	–23.49***	—	—
	(5.01)		
Election sequence (–)	–3.57***	–4.47***	–4.55***
	(0.59)	(1.04)	(1.12)
N	105	48	57
R^2	.471	.808	.683

Note: Entries are unstandardized regression coefficients, with panel-corrected standard errors (PCSEs) in parentheses.
*$p < .10$. **$p < .05$. ***$p < .01$.

Question #11: What do you conclude given your regression results? In other words, what do you want readers to take away from your table?

Hopefully, you found this an interesting exercise in both (1) thinking theoretically about why people hold the political beliefs they do, (2) how you run a regression to evaluate your theories, and (3) how you describe the results to others. Do let me know if you have any questions about any part of this process. This week is the culmination of the theoretical and empirical research we cover in this class. Congratulations for surviving this far! Next week we will reinforce what you have learned and discuss several ways that our data or theories can force us into other approaches besides the descriptive and inferential statistics we have discussed so far.

[2] Kostadinova, Tatiana, and Timothy J. Power. 2007. “Does Democratization Depress Participation? Voter Turnout in the Latin American and Eastern European Transitional Democracies.” Political Research Quarterly 60(3)” 363-377

Part 4: Data ninjas only (optional)

Time is short during the workshop, so I want to make sure that you take the time to run a theoretically motivated regression or two in Part 3 and summarise them in as much detail as you can.

If you are interested in a bit more practice, there are several additional steps that you may want to try your hand at.

First, we have focused on models with three variables. However, there are several additional variables that may be of interest in the dataset.

Step 10: Add additional variables to your regression model and interpret the results and how they change as you add more/different variables.

Second, you may want to try taking the same explanatory variables and see if they have similar or different effects to other political beliefs. For instance, if you modelled *anzus* you may want to switch with *usa*, or switch *scomo* for *albo*, *queen* for *republic*, etc.

Step 11: See if your theoretical/empirical model may also apply to different dependent variables.

Third and finally, we are often interested in the substantive significance of our models. Try plugging in different values of the independent variables into the regression equation ($Y = \alpha\ \beta_1 X +\ \beta_2 Z$). For instance, if you included age in your model see how the expected outcome changes as people age from the youngest age in the sample to the oldest or from the far left to the far right ideologically.

Step 12: Experiment with predicted values of your outcome given different values of your independent variables.