97/100

Group members: Kaiyue Wang, Kailu Wang

Part I: Presidential Voting

- 1. Generate a new variable
 - a. Generate a new variable high_income =1 if the person's family income is above the
 68th percentile (income=4 or income=5), high income=0 otherwise
- 2. Run a regression of voting Republican on your new high income variable use probit. In your table report the marginal effect at the mean of the independent variables! (see Class 20 slides and the programming resources below for how to do this)
 - a. Why are we using probit? Why do we report marginal effects and not the direct coefficients?

We use probit because we want to avoid the estimation of probabilities greater than 1 or less than 0. Because the interpretation on direct coefficients are not intuitive which we are actually interpret z-score. However, we can normally interpret the change by using marginal effects.

b. How do you interpret this marginal effect?

The person's family income that above the 68th percentile is associated with a 12.7 percentage point increase in the probability of voting to republican at the mean of high income.

3. Re-run this regression but add age, female, and race (use other race as the reference group) to your probit model. Comment on the effect of each of these variables on the probability of voting Republican and whether they are significant. Be sure that after your estimation you estimate the marginal effects at the means of the independent variable to report in the table. Age: Additional 1-year old increase is associated with 0.27 percentage point increase in probability voting to republican at the mean of high income.

Female: In general, female is 9.5 percentage point lower in probability to vote for republican than male at the mean of high income.

White: In general, white is 10.9 percentage point higher in probability to vote for republican than other race at the mean of high income.

Black: In general, black is 27.3 percentage point lower in probability to vote for republican than other race at the mean of high income.

All the variables are significant at 5% level, because all of them have at least two stars so based on our notes they are significant at 5% level.

4. Use this data to tell me something else interesting about the determinants of voting republican in the 2000 presidential election (make this a 3rd regression), and interpret your new marginal effect of interest for me.

South: In general, people who live in south is 14 percentage point higher in probability to vote for republican than the people that don't live in south at the mean of high income.

College: In general, attend college is 0.45 percentage point higher in probability to vote for republican than the people that not attend college at the mean of high income.

Some data that determinants of republican voting

	(1)	(2)	(3)
High income	0.127 ***	0.078 **	0.086**
	(0.031)	(0.03)	(0.033)
age		0.0027 **	0.0027**
		(0.0012)	(0.0013)
female		-0.095 ***	-0.098**
		(0.0301)	(0.030)
black		-0.273 ***	-0.28***
		(0.037)	(0.035)
white		0.109 ***	0.12**
		(0.04)	(0.04)
South			0.14***
			(0.036)
college			0.0045
			(0.033)

Constant			
(leave blank line			
above # of obs)			
Number of	1015	1015	1015
Observations			
Adj R ²	$[R^21]$	$[R^22]$	$[R^32]$

Notes: Standard errors in parenthesis under the marginal effects. stars on the coefficients to indicate statistical significance (*** for p<0.01, ** for p<0.05, * for p<0.1). rvote=1 if voted republican, =0 otherwise. high_income =1 if the person's family income is above the 68th percentile (income=4 or income=5), high_income=0 otherwise. High income, not attend college, not live in south, male and other race are omitted dummy variables.

Part II: Determinants of the State Suicide Rate

The following questions are based on a state level panel data on the suicide rates from 1990-2000 and was originally used in the paper "Are Mental Health Insurance Mandates Effective? Evidence from Suicides" *Health Economics* (2006) by Jonathan Klick and Sara Markowitz. The codebook is important and available to download with this assignment from Sakai.

Data that determinants of State suicide rate

	(1)	(2)	(3)	(4)
Unemployment	-0.26 **	0.24	0.17	0.105
rate	(0.12)	(0.12)	(0.12)	(0.18)
per capita state	-0.078***	-0.067 **	-0.053**	-0.045**
income	(0.014)	(0.027)	(0.023)	(0.025)
percent living in	-0.032**	0.025 **	0.22**	0.24**
rural areas	(0.015)	(0.078)	(0.077)	(0.076)
percent of	0.041	-0.12 **	-0.0898	-0.13**
population with a	(0.076)	(0.059)	(0.063)	(0.06003)
college degree				
percent of	0.29***	-0.0084	-0.036	-0.0195
uninsured	(0.051)	(0.056)	(0.0596)	(0.056)
Mormon			0.367	
			(0.33)	
Southern Baptist			-0.52	
			(0.32)	
Roman Catholic			-0.0297	
			(0.13)	
Mainline			0.238	
Protestant			(0.101)	
				0.0000000000000000000000000000000000000
Adult population				-0.0000024***
1.1				(0.00000035)
state population				-0.34
that is black				(0.25)
Constant				
Constant				

(leave blank line above # of obs)				
Number of	561	561	561	561
Observations				
Adj R ²	0.2435	0.8584	0.8606	0.864
State Fixed	No	Yes	Yes	Yes
Effects				
Year Fixed	No	Yes	Yes	Yes
Effects				

Notes: Standard errors in parenthesis under the marginal effects. stars on the coefficients to indicate statistical significance (*** for p<0.01, ** for p<0.05, * for p<0.1). adultsuic: number of suicides among adults aged 18-65. adultsuicrate: number of suicides amongst adults per 100,000 adults in the state. We fixed state and year effects in column 2,3,4.

- Run a regression of the adult suicide rate on percent unemployed, per capita state income, percent living in rural areas, percent of population with a college degree, and percent of uninsured
 - a. How do you interpret the coefficient on the unemployment rate?

An additional percent increase in unemployment rate is associated with a decrease of 0.26 percent of number in suicides amongst adults per 100,000 adults in the state.

b. Are you surprised by the sign of the coefficient?

Yes, because as unemployment rate increase the suicides rate should increase instead of decrease so it's surprised me.

- 2) Re-run the regression but include both state and year fixed effects. (see programming resources below and lecture notes from Class 25)
 - a. Which of your coefficients changed sign and significance from the regression in (1)?

Unemployment rate, percent living in rural areas, percent of population with a college degree, and percent of uninsured all change the sign. Unemployment rate change because in column (1) it has 2 stars but now it has no star.

Percent of population with a college degree change a lot because it became significant at 5% level.

Percent of uninsured also change because it became not significant at 5% level.

b. What are potential explanations for these changes?

Fixed effects are just a way of doing this even we have more than 2 time periods. It could help us to control for things that vary across year and state. Also, as we fixed state and year effects we could avoid omitted variable bias that arises from things that vary across time state.

- 3) Re-run your regression but now also include the available religion variables
 - a. How do you interpret the coefficient on percent southern Baptist?

An additional increase in Southern Baptist is associated with a decrease of 0.52 percent of number in suicides amongst adults per 100,000 adults in the state, controlling for state and year fixed effects.

- b. Are the religion variables jointly significant? How do you know? (what is the value of the statistic you used? And the p-value?)
- The religion variables are not jointly significant because according to P-value on F-statistic which is 0.13 that is larger than 0.05 so it's not jointly significant.
- 4) Use this dataset to tell me something else interesting about the suicide rate, be sure to include state and year fixed effects. Make this a 4th regression and interpret 1 or more

coefficients for me. Do you think that this effect you estimated is causal? Do we have more confidence in the causal interpretation than if we ran a cross-sectional regression that used data in a single year only? Why or why not?

I think this effect is not causal because we fixed the state and year affects so it's not causal. Also, we use randomized experiments which means error term cannot be correlated with independent variables.

We don't have more confidence in the causal interpretation than if we ran a crosssectional regression that used data in a single year only because we have higher probability in a single year only and we have a higher causality.

5) You used heteroscedasticity-robust standard errors, do you have reason to believe these standard errors may be incorrect? Why or why not? **Extra Credit:** rerun your regression in (3) with the appropriate standard errors and show me, did anything change?

We can think of reasons where the error terms might be correlated over time within a state. This means that the error terms may be autocorrelated which is always a problem with time-series data so these standard errors may be incorrect.