

Problem Set 5: Bivariate Regression

In this problem set, different explanations are considered, and the independent and dependent variables, direction of effect, potential rival hypothesis, intervening variables, and proper-form hypotheses are described for each. Next, two hypotheses based on data from the 2020 American National Election Study are considered. Prior to testing the hypotheses, the data is cleaned to remove missing values. For each hypothesis, bivariate regression is performed to test the relationship of the independent and dependent variables. The direction and magnitude of effect, and the strength of relationship between the variables, are also considered. Finally, the associated STATA code is attached in the Appendix.

Part 1: Variables and Research Design

Explanation 1

- a) For Explanation 1, the independent variable is the status of a hegemon. The dependent variable is the level of international trade.
- b) The direction of effect between the variables is positive, the stronger the hegemon the higher level of international trade.
- c) A rival hypothesis for Explanation 1 is the existence of any large country will influence international trade.
- d) Potential intervening variables include the amount of reserve currency held by the hegemon and the presence of trade disputes.
- e) In a comparison of levels of international trade, times with strong hegemons will be more likely to have high levels of international trade than times with declining or nonexistent hegemons will.

Explanation 2

- a) For Explanation 2, the independent variable is egg consumption. The dependent variable is a heart disease diagnosis.
- b) The direction of effect between the variables is positive, those that ate the greatest number of eggs per week were more likely to be diagnosed with heart disease.
- c) A rival hypothesis to Explanation 2 is that those who are predisposed to heart disease enjoy eggs more than people who are not predisposed to heart disease.
- d) A potential intervening variable is the rest of an individual's diet.

Bailey Williams
 Problem Set 5
 Due Date: July 23, 2025
 Word Count: 856
 Submitted: July 28, 2025

- e) In a comparison of individuals, those who consume more eggs will be more likely to be diagnosed with heart disease than will those who eat the least amount of eggs.

Part 2: Bivariate Regression Analysis

Hypothesis 1

Table 1: Bivariate Regression Analysis of Support of Government Spending on Social Security on the Function of Age

Source	SS	df	MS	Number of obs	=	7,897
Model	126.53	1	126.53	<i>F</i> (1, 7895)	=	125.51
Residual	7773.27	7,895	.98	<i>Prob > F</i>	=	0.00
Total	7899.79	7,896	1.00	<i>R-squared</i>	=	0.02
				<i>Adj R-squared</i>	=	0.02
				<i>Root MSE</i>	=	0.99

Support of Government Spending on Social Security	Coefficient	Std. err.	t	P>t	[95% conf. interval]	
Age (Consists of individual values for ages 18-79 and one value for 80 and older)	-.01	.00	-11.34	0.00	-0.01	-0.01
_cons	2.54	.04	71.99	0.00	2.47	2.61

- a) There is a negative direction of effect, which means that as age increases, support for increased government spending on Social Security increases. This is consistent with the hypothesis.
- b) The magnitude of effect is 0.01, meaning that for every year increase in age, support for government spending on Social Security increases by 0.1 points.
- c) The strength of the relationship, R-squared, is 0.02. This is a weak relationship, explaining 2 percent of the variation in support of Social Security funding as a function of age.
- d) The hypothesis is accepted because the p-value is less than 0.05. The p-value is displayed as 0.00, which means that while it is statistically nonzero, it is a very small number.

- e) As age increases, there is more support of increased government spending on social security. Since the p-value is 0.00, less than the critical threshold of 0.05, the null hypothesis that there is no relationship between age and support of government spending on social security is rejected.

Hypothesis 2

Table 2: Bivariate Regression Analysis of Ideologically Conservative Views on the Function of Household Income

<i>Source</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>Number of obs</i>	<i>=</i>	<i>6,573</i>
Model	.760612515	1	.760612515	<i>F(1, 6571)</i>	<i>=</i>	<i>0.27</i>
Residual	18352.0527	6,571	2.79288582	<i>Prob > F</i>	<i>=</i>	<i>0.60</i>
Total	18352.8133	6,572	2.79257659	<i>R-squared</i>	<i>=</i>	<i>0.00</i>
				<i>Adj R-squared</i>	<i>=</i>	<i>-0.00</i>
				<i>Root MSE</i>	<i>=</i>	<i>1.67</i>

<i>Political Ideology</i>	<i>Coefficient</i>	<i>Std. err.</i>	<i>t</i>	<i>P> t </i>	<i>[95% conf. interval]</i>	
Household Income	-0.00	.00	-0.52	0.60	-0.01	0.00
_cons	4.09	.04	93.95	0.00	4.00	4.17

- a) The direction of effect is negative, meaning that as household income increases individuals are more ideologically liberal than those with less household income. This is inconsistent with the hypothesis, which suggests the opposite.
- b) The magnitude of effect is 0.001, meaning that for every 0.1 unit increase in the household income variable, the individuals ideology becomes more liberal by 0.01 points.
- c) The strength of the relationship, R-squared, is 0.00, which means that there is a very weak relationship, explaining less than 1 percent of the variation in political ideology as a function of household income.
- d) The hypothesis is rejected because the p-value is 0.6, which is above the critical threshold of 0.05.
- e) As household income increases, there is more self-placement of liberal ideologies. Since the p-value is 0.06, the hypothesis is rejected and the null hypothesis that

Bailey Williams

Problem Set 5

Due Date: July 23, 2025

Word Count: 856

Submitted: July 28, 2025

there is no relationship between household income and political ideology is accepted.

Bailey Williams
Problem Set 5
Due Date: July 23, 2025
Word Count: 856
Submitted: July 28, 2025

Appendix

```
use "/Users/baileywilliams/Downloads/anes_timeseries_2020 (1).dta"  
  
replace V201507x = . if V201507x<0  
  
replace V201302x = . if V201302x<0  
  
regress V201302x V201507x  
  
replace V201617x = . if V201617x<0  
  
replace V201200 = . if V201200<0  
  
replace V201200 = . if V201200>10  
  
regress V201200 V201617x  
  
save "/Users/baileywilliams/Downloads/anes_timeseries_2020 (1).dta", replace
```