# Jeff Davis

# POLS 418 Quantitative Methods

Problem Set 5

Due June, 17th 2020

Submitted June, 17th 2020

Word Count: 984

*Problem Set #5:*

*Bivariate and Multivariate Regression*

Please complete each task listed below. You should type your problem set and turn in a hard copy to me. You do not need to include a cover page. I expect, however, that your first page will include your full name, the problem set number, the due date, and the date you submit the work. The final page(s) of your assignment ***must*** include the STATA code you used, otherwise it is incomplete, and it will receive the grade of “I”. Likewise, I will grade your assignment “I” if you use screenshots from STATA instead of making your own tables in MS Word (note: if you want to do section 3 by hand that is fine).

In this problem set, we will use the NES2016.dta dataset (the same dataset we used for Problem Set 4) to test two hypotheses about an individual’s orientations toward government guaranteed jobs and incomes and Christian Fundamentalists. The data comes from the National Election Survey (NES) of 2016, which asked respondents about their opinions on a range of topics.

Please provide all corresponding tables and figures to support your analysis.

1. Please conduct a bivariate regression analysis (i.e. one dependent variable, one independent variable), then answer questions (a) through (d) below. Make sure you have properly dealt with all missing/inappropriate values before conducting the regression analysis.

*Ha1*: In comparing individuals, those with more education will be less supportive of government guaranteed jobs and income (V161189), than those with less education.

Table. 1 Regression Analysis of Education and View Government Guaranteed jobs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Opposition to Government  Guaranteed Jobs and Income | Coef. | Std. Err. | t | P>t | [95% Conf. | Interval] |
| education | .033 | .012 | 2.59 | 0.010 | .008 | .058 |
| \_cons | 3.909 | .148 | 26.32 | 0.000 | 3.618 | 4.20 |

*R-squared=.0018*

1. **What is the direction of effect? Is it consistent with or contrary to the hypothesis?**

The direction of the relationship stated in the hypothesis is negative. But the direction of the relationship in the regression analysis is positive. However, the regression is consistent with the hypothesis. The difference in the direction of the effect is due to the coding of the dependent variable. V161189—the dependent variable—is a 7.0-point scale survey question that asks to rank support for government guaranteed jobs and income with support rated at 1.0 and opposition rated at 7.0. Therefore, the variable is also a measure of opposition. As a person’s education increases, they are more likely to oppose government guaranteed jobs.

1. **What is the magnitude of effect, if any?**

The magnitude of the effect is .03. For each one unit increase in education, there is a .03 increase in the expected value on the dependent variable—opposition to government guaranteed jobs.[[1]](#footnote-1)

1. **What is the strength of the relationship (i.e. the association), if any?**

The strength of the relationship is relatively weak. The R-squared value is a measure of the strength of the relationship. A regression analysis reported an R-squared value of .0018. This means that education has only a .18% effect on the dependent variable, leaving 99.82% of the relationship unexplained.

1. **Do you accept or reject the hypothesis? How confident are you (i.e. what is the p-value)?**

I accept the claim in the hypothesis that an increase in education is associated with less support (or increased opposition) to government guaranteed jobs. A regression analysis reported a p-value of .010, below the accepted .05 threshold for rejecting the null hypothesis. However, both the low magnitude of the effect and the t-score, which is only slightly above the accepted threshold, renders the relationship substantively insignificant.

1. Select a variable from the 2016 ANES you feel might affect attitudes towards government guaranteed jobs and income. Add this additional variable to the previous regression model that included ‘education’ as an independent variable (note: still include education, you are adding a second independent variable). Make sure you have properly dealt with all missing/inappropriate values before conducting the regression analysis.

Table. 2 Regression Analysis of Education and Part ID with View on Government Guaranteed Jobs and Income

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Opposition to Government  Guaranteed Jobs and Income | Coef. | Std. Err. | t-statistic | P>t | [95% Conf. | Interval] |
| Education | .006 | .0142 | 0.44 | 0.661 | -.021 | .0341 |
| Party ID | 1.836 | .065 | 27.98 | 0.000 | 1.708 | 1.965 |
| \_cons | 1.529 | .191 | 7.97 | 0.000 | 1.153 | 1.905 |

*R-squared=0.2478*

1. **What is the direction of effect? Is it consistent with or contrary to the hypothesis?**

When controlling for party identification—Democrat or Republican—the direction of effect for education was still positive. The direction of effect for party ID was positive. Party ID was coded 1-Democrat and 2-Republican. A positive effect suggests that a republican is more likely to oppose government guaranteed jobs and income.

1. **What is the magnitude of effect, if any?**

The magnitude of effect for party ID was 1.836, meaning that each 1-unit change in party ID was associated with a 1.836 increase in the expected value of the dependent variable—opposition to government guaranteed jobs and income. Compared to the magnitude of effect for education, party ID had a considerably larger effect on the dependent variable.

1. **What is the strength of the relationship, if any?**

When controlling for party ID, the strength of the relationship was still relatively weak. Although the combined effect of both independent variables, substantially increased the strength of the relationship compared to the original bivariate analysis. For example, in the first regression the R-squared value was .0018. In the multivariate analysis, the R-squared value was 0.2478. This means that in the multivariate analysis the independent variables explained 24.78% of the variation on the dependent variable. However, a comparison of the magnitudes suggest that party ID had the larger effect, even with taking into consideration the structure of the measurements.

1. **Do you accept or reject the hypothesis? How confident are you?**

Based on the p-value and the t-statistic for education in the multivariate analysis, I reject the hypothesis that people with more education are less likely to support (or more likely to oppose) government guaranteed jobs and income. Although the p-value and t-statistic for education met the accepted thresholds in the bivariate analysis, they did not meet those standards when controlling for party ID. In the multivariate analysis the p-value and t-statistic for education was 0.661 and 0.44 respectively.

1. **Did adding this third variable alter the effect of ‘education’?**

Yes. Adding a third variable reduced the effect of education on the dependent variable. For example, in the bivariate analysis the magnitude of education was .03. But in the multivariate analysis, the magnitude of education was .006.

1. Please conduct a bivariate regression analysis (i.e. one dependent variable, one independent variable), then answer questions (a) through (d) below. Make sure you have properly dealt with all missing/inappropriate values before conducting the regression analysis.

*Ha1*: In comparing individuals, those with more income (income) will have a less favorable view of Christian Fundamentalists (V162095) those with less income.

Table. 1 Regression Analysis of Income and Christian Fundamentalist Feeling Thermometer

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Christian Fundamentalist Feeling Thermometer | Coef. | Std. Err. | t | P>t | [95% Conf. | Interval] |
| income | -.5476 | .058 | -9.32 | 0.000 | -.662 | -.4324 |
| \_cons | 58.683 | 1.027 | 57.13 | 0.000 | 56.669 | 60.697 |

R-squared=0.0251

1. **What is the direction of effect? Is it consistent with or contrary to the hypothesis?**

The direction is negative. As income increases, a person tends to have a less favorable view of Christian Fundamentalists. The results of the regression analysis are consistent with the stated hypothesis.

1. **What is the magnitude of effect, if any?**

The magnitude of the effect is approximately -.55. This means that for each one-unit change in income, there is a .55 unit decrease on the Christian Fundamentalist feeling thermometer.

1. **What is the strength of the relationship (i.e. the association), if any?**

The strength of the relationship is relatively weak. The R-squared value was approximately .03, meaning that income had only a 3% effect on the dependent variable.

1. **Do you accept or reject the hypothesis? How confident are you (i.e. what is the p-value)?**

Based on the p-value and the t-score for income, I accept the hypothesis that an increase in income is negatively associated with feelings towards Christian fundamentalists. The p-value on the regression analysis was less than .0001, well below the accepted threshold of .005. Similarly, the t-score was -9.32, well above the accepted threshold of 2.0.

1. Select a variable from the 2016 ANES you feel might affect attitudes towards Christian Fundamentalists. Add this additional variable to the previous regression model that included ‘income’ as an independent variable (note: still include income, you are adding a second independent variable). Make sure you have properly dealt with all missing/inappropriate values before conducting the regression analysis.

Table. 1 Regression Analysis of Income and Conservative Feeling Thermometer on Christian Fundamentalist Feeling Thermometer

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Christian Feeling Thermometer | Coef. | Std. Err. | t | P>t | [95% Conf. | Interval] |
| Income | -.540 | .0489 | -11.04 | 0.000 | -.636 | -.444 |
| Conservative feeling thermometer | .609 | .0156 | 38.82 | 0.000 | .578 | .640 |
| \_cons | 24.458 | 1.228 | 19.91 | 0.000 | 22.050 | 26.867 |

*R-squared=0.3271*

1. **What is the direction of effect? Is it consistent with or contrary to the hypothesis?**

When controlling for feelings towards conservatives, the direction of effect for income on feelings towards Christian fundamentalist was still negative. The direction of effect for feelings towards conservatives on feelings towards Christian fundamentalist was positive. When adding a control, the direction of effect of income was still consistent with the original hypothesis.

1. **What is the magnitude of effect, if any?**

The magnitude of the effect is approximately .609, meaning that for everyone unit increase in the conservative feeling thermometer, there was an increase of .609 on the Christian fundamentalist feeling thermometer. Comparatively, the magnitude of effect for income was greater than that of the conservative feeling thermometer, even though the absolute values suggest the opposite.

1. **What is the strength of the relationship, if any?**

When controlling for feelings towards conservatives, the strength of the relationship was still relatively weak. However, the strength of the relationship improved substantially when adding the control. For example, in the first regression the R-squared was 0.0251. In the multivariate analysis the R-squared was .3271. This means that in the multivariate analysis the combined effect of the independent variables explained 67.29% of the variation in the dependent variable. A comparison of the magnitudes suggest that income had the larger effect.

1. **Do you accept or reject the hypothesis? How confident are you?**

Based on the p-values and t-scores for both independent variables, I accept the alternative hypothesis that those with more income are more likely to have less favorable views towards Christian fundamentalists. Both p-values are less than .0001, well below the accepted .05 threshold. Similarly, the t-scores were -11.04 and 38.82, well above the accepted 2.0 threshold

1. **Did adding this third variable alter the effect of ‘income’?**

Yes. Adding the third variable altered the effect of income. Although the change appears effect negligible. The original magnitude of income was -.547 while the magnitude on the subsequent regression was -.540.

1. Political scientists love to watch *The Simpsons*, so much so that some will watch it for several hours a day. The following figures are the average number of hours per day that seven political scientists spend watching re-runs of *The Simpsons*.

1 5 .5 1 9 8 11 3 0

Make the following calculations by hand and show your work.

1. **What is the sum of squares for the number of hours these political scientists spent watching *The Simpsons*?**

|  |  |  |
| --- | --- | --- |
| **X** | X-X̅ | (X-X̅)2 |
| 1 | 1- 4.27777778= -3.277777778 | 10.7438272 |
| 5 | 5- 4.27777778= 0.722222222 | 0.52160494 |
| .5 | .5- 4.27777778= -3.777777778 | 14.2716049 |
| 1 | 1- 4.27777778= -3.277777778 | 10.7438272 |
| 9 | 9- 4.27777778= 4.722222222 | 22.2993827 |
| 8 | 8- 4.27777778= 3.722222222 | 13.8549383 |
| 11 | 11-4.27777778= 6.722222222 | 45.1882716 |
| 3 | 3 - 4.27777778= -1.277777778 | 1.63271605 |
| 0 | 0- 4.27777778= -4.277777778 | 18.2993827 |

X̅ = 4.27777778 **Σ = 137.555556**

1. **What is the variance?**

|  |  |  |
| --- | --- | --- |
| **X** | X-X̅ | (X-X̅)2 |
| 1 | 1- 4.27777778= -3.277777778 | 10.7438272 |
| 5 | 5- 4.27777778= 0.722222222 | 0.52160494 |
| .5 | .5- 4.27777778= -3.777777778 | 14.2716049 |
| 1 | 1- 4.27777778= -3.277777778 | 10.7438272 |
| 9 | 9- 4.27777778= 4.722222222 | 22.2993827 |
| 8 | 8- 4.27777778= 3.722222222 | 13.8549383 |
| 11 | 11-4.27777778= 6.722222222 | 45.1882716 |
| 3 | 3 - 4.27777778= -1.277777778 | 1.63271605 |
| 0 | 0- 4.27777778= -4.277777778 | 18.2993827 |

Σ = 137.555556/n-1 = 17.1944444

**s2 = 17.1944444**

1. **What is the standard deviation?**

|  |  |  |
| --- | --- | --- |
| **X** | X-X̅ | (X-X̅)2 |
| 1 | 1- 4.27777778= -3.277777778 | 10.7438272 |
| 5 | 5- 4.27777778= 0.722222222 | 0.52160494 |
| .5 | .5- 4.27777778= -3.777777778 | 14.2716049 |
| 1 | 1- 4.27777778= -3.277777778 | 10.7438272 |
| 9 | 9- 4.27777778= 4.722222222 | 22.2993827 |
| 8 | 8- 4.27777778= 3.722222222 | 13.8549383 |
| 11 | 11-4.27777778= 6.722222222 | 45.1882716 |
| 3 | 3 - 4.27777778= -1.277777778 | 1.63271605 |
| 0 | 0- 4.27777778= -4.277777778 | 18.2993827 |

Σ = 137.555556/n-1 = 17.1944444

(√) 17.1944444

**s= 4.14661843**

1. **The *z* scores**

z = X- X̅

S

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | X-X̅ | **X- X̅**  **S** | **Z-Scores** |
| 1 | 1- 4.27777778= -3.277777778 | -3.277777778  4.14661843 | -0.79 |
| 5 | 5- 4.27777778= 0.722222222 | 0.722222222  4.14661843 | 0.17 |
| .5 | .5- 4.27777778= -3.777777778 | -3.777777778  4.14661843 | -0.91 |
| 1 | 1- 4.27777778= -3.277777778 | -3.277777778  4.14661843 | -0.79 |
| 9 | 9- 4.27777778= 4.722222222 | 4.722222222  4.14661843 | 1.14 |
| 8 | 8- 4.27777778= 3.722222222 | 3.722222222  4.14661843 | 0.90 |
| 11 | 11-4.27777778= 6.722222222 | 6.722222222  4.14661843 | 1.62 |
| 3 | 3 - 4.27777778= -1.277777778 | -1.277777778  4.14661843 | -0.31 |
| 0 | 0- 4.27777778= -4.277777778 | -4.277777778  4.14661843 | -1.03 |

1. **The mean of the *z* scores**

0 (see uploaded excel)

1. **The standard deviation of the *z* scores**

1 (see uploaded excel)

**Appendix A**

**Figure 1. STATA code for problem set 5**

. tab education

. tab V161189

. replace V161189 = . if V161189<1

. replace V161189 = . if V161189>7

. regress V161189 education

. tab V161155

. replace V161155 = . if V161155<1

. replace V161155 = . if V161155>2

. regress V161189 education V161155

. tab income

. tab V162095

. replace V162095 = . if V162095<0

. replace V162095 = . if V162095>100

. regress V162095 income

. tab V162101

. replace V162101 = . if V162101>100

. replace V162101 = . if V162101<0

. regress V162095 income V162101

1. Education is rated on a 16.0-point scale where 1.0 is less than 1st grade and 16.0 is a doctorate degree. [↑](#footnote-ref-1)