

Practice with Regression Analysis

Preliminaries

Download the EViews file “school.wf1” from the course web site (under the Assignments section). In the file, there is a total of five variables:

| | |
|-------------|--|
| DRUGS: | Index of how much drug selling there is in the neighborhood |
| ENROLLMENT: | Enrollment of the student's school |
| MALE: | A “dummy variable” that takes the value of 1 if the student is a male, 0 if the student is a female |
| MATH87: | Math test score in 8th grade in standard deviations from the mean |
| MATH91: | Math test score in 12th grade in standard deviations from the mean |
| SES: | Socio-economic status = a combination of parent's education, income, and goods in the house (SES is measured in standard deviations from the mean) |
| URBAN: | % of people in school's zip code that live in an urban area |

Suppose that you are participating in a project that studies math proficiency at high schools. Your project asks the question: What are the important factors determining student's math test scores? Furthermore, what types of policies should school administrators implement to achieve improved math proficiency at high schools?

1. Which of the variables above is the dependent variable?
2. Which are the explanatory variables? Provide a one sentence explanation for how and why each of these variables affects the dependent variable you identified in question 1.

Importance of Scale

Run a simple regression of ENROLLMENT on the dependent variable. You will see the coefficient for ENROLLMENT is very small.

3. Create a table for with the output from the regression above. What is the coefficient for ENROLLMENT? Does school enrollment have a statistically significant effect on the dependent variable? Explain how you know.

Create a new variable called “ENROLL” that measures the enrollment of the student's school per 100 students. Now, run a regression of “ENROLL” on the dependent variable.

4. Create a table for with the output from the regression above. What is the coefficient for ENROLL? Does school enrollment have a statistically significant effect on the dependent variable? Is this result different from the one in question 3? Explain.
5. Why might a researcher use ENROLL as an explanatory variable instead of ENROLLMENT?

Regressions

A natural place to start is to run a “kitchen sink” regression that includes all of the explanatory variables that are available. Run a regression of the dependent variable on all of the explanatory variables, except for the “MALE” variable.

6. Create a table for with the output from the regression above.
7. Which explanatory variables have a significant effect on the dependent variable? Which ones do not?
8. Based on your regression results, if a student's 8th grade math test score increases by one standard deviation, how much will his/her 12th grade math test score increase?
9. Based on your regression results, if a student's socio-economic status improves by one (as measured by the index), what will happen to his/her 12th grade math test score?

Run another regression that includes only those explanatory variables that were significant in the regression above.

10. Create a table for with the output from the regression above.
11. Using the results from this regression, repeat the exercises from question 8 and 9.

Dummy Variables

Suppose you are interested in whether the difference in test scores is statistically different for males vs. females.

12. What is the average 12th grade math test score for all students? What is the average for males? For females?
13. Is the average test score for females statistically different from that of males? Write out a hypothesis test and report the t-statistic and p-value for this test.

If you believe there is a fundamental difference between the performance of males vs. females, you should include a “dummy variable” in your regression. This allows you to account for an average different between males and females in terms of math test scores that are not accounted for by the other variables in your data set.

14. Create a table with the results from a regression you did in question 10, including the MALE variable.
15. Is there a statistically significant different between math test scores for males vs. females? How do you rectify this with your answer to question 13?

Interpreting the Results

16. How can school administrators improve high school math proficiency? Are the variables that are statistically significant things that administrators have power over? Explain in four to five sentences.

*****Note on how to submit answers:**

Your answers should be presented in a typed assignment with appropriate tables and graphs labeled. You will not receive full credit if tables and graphs are not formatted in a way that makes them clear to a generic reader. Please see the Presentation of Data supplement for details and examples.