# Problem Set #6 : LDV Models and Regression in R

Economics 435: Quantitative Methods

Fall 2011

## 1 Public sector unionization and size of government: Part III

We continue our analysis of public sector unionization and the size of government, by estimating some regressions.

**a**) Fill in Tables 2 and 3 in the paper. You will find the command `lm` useful.

Just to be clear: each column in these tables refers to a separate regression. The text of the paper describes which regression is being estimated in each column. When reporting regression coefficients, we will follow the convention of reporting standard errors in parentheses below each coefficient, and of indicating statistical significance with stars. Three stars mean the coefficient is significant at 1%, two stars mean it is significant at 5%, and one star means it is significant at 10%. If you are unsure about how the table is supposed to be laid out, show me what you have and I'll give you advice.

**b**) Fill in Figure 3 in the paper. You will find the commands `supsmu` and `line` or `abline` useful.

## 2 Question 17.2 from the textbook

Let *grad* be a dummy variable for whether a student-athlete at a large university graduates in five years. Let $hsGPA$ and $SAT$ be high school GPA and SAT exam score respectively. Let *study* be the number of hours spent per week in an organized study hall. Suppose that, using data on 420 student-athletes, the following logit model is obtained:

$$\hat{P}(grad = 1|hsGPA, SAT, study) = \Lambda(-1.17 + 0.24hsGPA + 0.00058SAT + 0.73study)$$

where $\Lambda(z) = \frac{e^z}{1+e^z}$ is the logit function. Holding $hsGPA$ fixed at 3.0 and $SAT$ fixed at 1200, compute the estimated difference in the graduation probability for someone who spent 10 hours per week in study hall and someone who spent 5 hours per week.