# Additional Tutorial 1

# The Simple Linear Regression Model

### Problem 1 (Wooldridge, Problem 2.4)

Suppose you are interested in estimating the effect of hours spent in an SAT preparation course (*hours*) on total SAT score (*sat*). The population is all college-bound high school seniors for a particular year.

- (i) Suppose you are given a grant to run a controlled experiment. Explain how you would structure the experiment in order to estimate the causal effect of hours on sat.
- (ii) Consider the more realistic case where students choose how much time to spend in a preparation course, and you can only randomly sample sat and hours from the population. Write the population model as

$$sat = \beta_0 + \beta_1 hours = u$$

where, as usual in a model with an intercept, we can assume E(u) = 0. List at least two factors contained in u. Are these likely to have positive or negative correlation with hours?

- (iii) In the equation from part (ii), what should be the sign of  $\beta_1$  if the preparation course is effective?
- (iv) In the equation from part (ii), what is the interpretation of  $\beta_0$ ?

#### Problem 2 (Wooldridge, Problem 2.5)

Consider the savings function

$$sav = \beta_0 + \beta_1 inc + u, u = \sqrt{inc} \times e$$

- (i) Show that E(u|inc) = 0, so that the key zero conditional mean assumption (Assumption SLR.4) is satisfied. [Hint: If e is independent of inc, then E(e|inc) = E(e).]
- (ii) Show that  $\operatorname{Var}(u|inc) = \sigma_e^2 inc$ , so that the homoskedasticity Assumption SLR.5 is violated. In particular, the variance of sav increases with *inc*. [Hint:  $\operatorname{Var}(e|inc) = \operatorname{Var}(e)$ , if e and *inc* are independent.]
- (iii) Provide a discussion that supports the assumption that the variance of savings increases with family income.

#### Problem 3 (Wooldridge, Problem 2.7)

Using data from 1988 for houses sold in Andover, Massachusetts, from Kiel and McClain (1995), the following equation relates housing price (price) to the distance from a recently built garbage incinerator (dist):

$$log(price) = 9.40 + 0.312log(dist)$$
  
 $n = 135, R^2 = 0.162$ 

- (i) Interpret the coefficient on  $\log(dist)$ . Is the sign of this estimate what you expect it to be?
- (ii) Do you think simple regression provides an unbiased estimator of the ceteris paribus elasticity of price with respect to dist? (Think about the city's decision on where to put the incinerator.)
- (iii) What other factors about a house affect its price? Might these be correlated with distance from the incinerator?

## Problem 4 (Wooldridge, Problem 2.9)

In the linear consumption function

 $\widehat{cons} = \beta_0 + \beta_1 inc,$ 

the (estimated) marginal propensity to consume (MPC) out of income is simply the slope,  $\hat{\beta}_1$  while the average propensity to consume (APC) is  $\widehat{cons}/inc = \hat{\beta}_0/inc + \hat{\beta}_1$ . Using observations for 100 families on annual income and consumption (both measured in dollars), the following equation is obtained:

 $\widehat{cons} = -124.84 + 0.853 inc$ 

$$n = 100, R^2 = 0.692.$$

- (i) Interpret the intercept in this equation, and comment on its sign and magnitude.
- (ii) What is the predicted consumption when family income is \$30,000?
- (iii) With *inc* on the x-axis, draw a graph of the estimated MPC and APC.

# Problem 5 (Wooldridge, Problem 2.11)

The data set BWGHT.RAW contains data on births to women in the United States. Two variables of interest are the dependent variable, infant birth weight in ounces (bwght), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (cigs). The following simple regression was estimated using data on n = 1388 births:

$$b\widehat{wght} = 119.77 - 0.514 cigs$$

- (i) What is the predicted birth weight when cigs = 0? What about when cigs = 20 (one pack per day)? Comment on the difference.
- (ii) Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.
- (iii) To predict a birth weight of 125 ounces, what would *cigs* have to be? Comment.
- (iv) The proportion of women in the sample who do not smoke while pregnant is about .85. Does this help reconcile your finding from part (iii)?