

HW week 10

w203: Statistics for Data Science

Tako Hisada

11/13/2017

1. Recall that the slope coefficient in a simple regression of Y_i on X_i can be expressed as,

$$\beta_1 = \frac{\text{cov}(X_i, Y_i)}{\text{var}(X_i)}$$

Suppose that you were to add a random variable, M_i , representing measurement error, to each X_i . You may assume that M_i is uncorrelated with both X_i and Y_i . You then run a regression of Y_i on $X_i + M_i$ instead of on X_i . Does the measurement error increase or decrease your slope coefficient?

The file `bwght.RData` contains data from the 1988 National Health Interview Survey. It was used by J. Mullahy for a 1997 paper ("Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *Review of Economics and Statistics* 79, 596-593.) and provided by Wooldridge. You will use this data to examine the relationship between cigarette smoking and a child's birthweight.

```
load("bwght.RData")
```

1. Examine the dependent variable, infant birth weight in ounces (`bwght`) and the independent variable, the number of cigarettes smoked by the mother each day during pregnancy (`cigs`).
2. Fit a linear model that predicts `bwght` as a function of `cigs`. Superimpose your regression line on a scatterplot of your variables.
3. Examine the coefficients of your fitted model. Explain, in particular, how to interpret the slope coefficient on `cigs`. Is it practically significant?
4. Write down the two moment conditions for this regression. Use R to verify that they hold for your fitted model.
5. Does this simple regression capture a causal relationship between smoking and birthweight? Explain why or why not.
6. Does your scatterplot show evidence of measurement error in `cigs`? If so, what does this say about the true relationship between cigarettes and birthweight?
7. Using your coefficients, what is the predicted birthweight when `cigs` is 0? When `cigs` is 20?
8. Use R's `predict` function to verify your previous answers. You may insert your linear model object into the command below.

```
predict(your_lm_object , data.frame(cigs = c(0, 20) ) )
```

9. To predict a birthweight of 100 ounces, what would `cigs` have to be?