İKT 452/554 Applied Microeconometrics

Homework 1:

Question 1: Use the data in WAGE2.RAW (Wooldridge datasets) to estimate a simple regression explaining monthly salary (*wage*) in terms of KWW score (a measure of ability, the score on the Knowledge of the World of Work test).

(i) Find the average salary and average KWW in the sample. What is the standard deviation of KWW?

(ii) Estimate a simple regression model where a one-point increase in *KWW* changes *wage* by a constant dollar amount. Use this model to find the predicted increase in wage for an increase in *KWW* of 7 points. Does *KWW* explain most of the variation in *wage*?

(iii) Now estimate a model where each one-point increase in *KWW* has the same percentage effect on *wage*. If *KWW* increases by 7 points, what is the approximate percentage increase in predicted *wage*?

Question 2: Use the data in WAGE1.RAW (Wooldridge datasets).

a) What are the shares of those with educ<=12, those with 13<=educ<=15, and those with educ>=16?

b) Define education dummies (Hint: Define one dummy for those who have 13 to 15 years of education (some college education) and another dummy for those who have 16 or more years of education (completed college education).)

b) Estimate a model that shows you the percentage effect of moving from the base category of education to a higher category of education on *wage.* (What is the base category here?) In addition to education dummies, include dummy variables for nonwhite, female, married, smsa and quadratic functions of exper and tenure as control variables. Interpret the coefficient estimates of the education dummies.

c) Test in Stata if the wage return to completing college education is the same as the wage return to getting only some college education. Comment on the output of the test.

Question 3: Use the data in WAGE1.RAW (Wooldridge datasets).

a) Estimate a model that expresses log(wage) as a function of educ, exper, nonwhite, female, smsa, and numdep. Compare the coefficient estimates of education and experience.

b) Estimate the beta coefficients (standardized coefficients). Compare the standardized coefficient estimates of education and experience.

*You can manually standardize all variables and run OLS using those variables.
*to manually standardize
summarize VARNAME

```
local a = r(mean)
local b = r(sd)
/* check: */
disp `a', `b'
gen std_VARNAME=(VARNAME -`a')/`b'
```

Question 4: Use the data in HPRICE3.RAW (Wooldridge datasets). The data are for houses that were sold in North Andover, MA. Use the observations for year 1981, which was the year construction began on a local garbage incinerator.

a) We want to study the effects of the incineartor location on house prices. Estimate the following regression model:

$$\log(price) = \beta_0 + \beta_1 \log(dist) + u \,,$$

where price is house price in dollars and dist is the distance of the house from the incinerator, measured in feet. What sign do you expect β_1 to take, assuming that the presence of the incinerator depresses house prices? Interpret your estimates.

b) To the simple regression model in part (a), add the variables log(*inst*), log(*area*), log(*land*), *rooms*, *baths*, and *age*, where *inst* is distance from the home to the interstate, *area* is square footage of the house, *land* is the lot size in square feet, *rooms* is total number of rooms, *baths* is number of bathrooms, and *age* is age of the house in years. Now what do you conclude about the effects of the incinerator? Explain why (a) and (b) give conflicting results.

(c) Add $[\log(inst)]^2$ to the model from part (b). Now what happens? What do you conclude about the importance of functional form? What is the best distance to the interstate?

(d) Is the square of log(*dist*) significant when you add it to the model from part (c)?

Question 5:

a) $\sum_{i=1}^{n}(x_i - \bar{x}) = ?$

b) Show that $\sum_{i=1}^{n} x_i(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})$

c) Show that $\sum_{i=1}^{n} y_i(x_i - \bar{x}) = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$