University of California, Santa Barbara **Olivier Deschenes**
Department of Economics **Winter 2018**

ESM 296
Individual Assignment 1

Due in class 01/31/18

Some of these exercises are taken from the Stock and Watson textbook.

Question 1:

Sir Francis Galton, a cousin of James Darwin, examined the relationship between the height of children and their parents towards the end of the 19th century. It is from this study that the name "regression" originated. You decide to update his findings by collecting data from 110 college students, and estimate the following relationship:

$$\underset{}{\widehat{Studenth}} = \underset{(7.2)}{19.6} + \underset{(0.10)}{0.73} \times Midparh, \quad R^2 = 0.45$$

where *Studenth* is the height of students in inches, and *Midparh* is the average of the parental heights. (Following Galton's methodology, both variables were adjusted so that the average female height was equal to the average male height.). The numbers in parenthesis are the heteroskedasticity-robust standard errors.

(a) Interpret the estimated coefficients.

(b) What is the meaning of the regression R^2?

(c) What is the prediction for the height of a child whose parents have an average height of 70.06 inches?

(d) Given the positive intercept and the fact that the slope lies between zero and one, what can you say about the height of students who have quite tall parents? Who have quite short parents?

(e) Test for the statistical significance of the slope coefficient.

(f) If children, on average, were expected to be of the same height as their parents, then this would imply two hypotheses, one for the slope and one for the intercept.

  (i) What should the null hypothesis be for the intercept? Calculate the relevant *t*-statistic and carry out the hypothesis test at the 1% level.

(ii) What should the null hypothesis be for the slope? Calculate the relevant *t*-statistic and carry out the hypothesis test at the 5% level.

Question 2

The data for this question contain information on the reported value and characteristics of houses in the Boston area we used in class (note that the file for the assignment has more observations). The STATA data file "HPRICE2.dta" is available on the class website. The same file also available in spreadsheet format "HPRICE2.csv".

Consider the following linear regression model for the price and characteristics of houses:

$$\mathrm{Price}_{\mathrm{i}} = \beta_0 + \beta_1 NOx_i + \beta_2 Rooms + \beta_3 STratio_i + u_i$$

Where *price* is the value of the house, NOx is a measure of NOx concentration in the Census track (in parts per 100 million), *Rooms* is the number of rooms in the house, and STratio is the student-teacher ratio in the nearest school.

(a) What is the effect of adding an additional room on the house price, holding NOx concentrations and student-teacher ratio constant?

(b) What is the estimated effect on house values of reducing NOx concentrations by 2.5 parts per 100 million?

(c) What percent of the variation in house values is explained by NOx concentrations, number of rooms, and student-teacher ratio? What percent of the variation in house values is explained by NOx concentrations alone?

(d) What is the predicted value of a house with NOx concentrations of 6, with 7 rooms, and a student-teacher ratio of 20? The actual price for that house was $20,000. Did the buyer overpay for this house?

(e) Test the null hypothesis that β_1 and β_2 jointly equal zero.

(f) Test the null hypothesis that $\beta_1 = \beta_3$ against the two-sided alternative.