Final Exam

625.603: Statistical Methods and Data Analysis

Please present all responses and supporting work in a single document. Present all relevant code and output. If you choose to write your solutions in your code, please clearly indicate solutions (i.e., use colors, boxing, etc.).

Data Set: **CPS2015.dta**
Description: The datafile contains data for 2015 for full-time workers with a high school diploma or B.A./B.S. as their highest degree. See the pdf attachment for an overview of the data and variable descriptions. In this exercise, you will investigate the relationship between a worker's age and earnings. (Generally, older workers have more job experience, leading to higher productivity and higher earnings.)[1]

a. Run a regression of average hourly earnings (*AHE*) on age (*Age*), gender (*Female*), and education (*Bachelor*). If age increases from 25 to 26, how are earnings expected to change? If age increases from 33 to 34, how are earnings expected to change?
b. Run a regression of the logarithm of average hourly earnings, ln(*AHE*), on *Age*, *Female*, and *Bachelor*. If age increases from 25 to 26, how are earnings expected to change? If age increases from 33 to 34, how are earnings expected to change?
c. Run a regression of the logarithm of average hourly earnings, ln(*AHE*), on ln(*Age*), *Female*, and *Bachelor*. If age increases from 25 to 26, how are earnings expected to change? If age increases from 33 to 34, how are earnings expected to change?
d. Run a regression of the logarithm of average hourly earnings, ln(*AHE*), on *Age*, Age^2, *Female*, and *Bachelor*. If age increases from 25 to 26, how are earnings expected to change? If age increases from 33 to 34, how are earnings expected to change?
e. Do you prefer the regression in (c) to the regression in (b)? Explain.
f. Do you prefer the regression in (d) to the regression in (b)? Explain.
g. Do you prefer the regression in (d) to the regression in (c)? Explain.
h. Run a regression of ln(*AHE*), on *Age*, Age^2, *Female*, *Bachelor*, and the interaction term *Female*Bachelor*. What does the coefficient on the interaction term measure? Alexis is a 30-year-old female with a bachelor's degree. What does the regression predict for her value of ln(*AHE*)? Jane is a 30-year-old female with a high school degree. What does the regression predict for her value of ln(*AHE*)? What is the predicted difference between Alexis's and Jane's earnings? Bob is a 30-year-old male with a bachelor's degree. What does the regression predict for his value of ln(*AHE*)? Jim is a 30-year-old

---

male with a high school degree. What does the regression predict for his value of ln(*AHE*)? What is the predicted difference between Bob's and Jim's earnings?

i. Is the effect of age on earnings different for men than for women? Specify and estimate a regression that you can use to answer this question.

j. Is the effect of age on earnings different for high school graduates than for college graduates? Specify and estimate a regression that you can use to answer this question.

k. After running all these regressions, summarize the effect of age on earnings for young workers.

Extra Credit (5 points):

In a few sentences, describe something you learned/discovered through this assignment.