

R computing for Business Data Analytics

Homework 5 (Due date: December 16, 2014)

Please e-mail your homework (.pdf) and the associated R code (.R) to hchuang.om@gmail.com.

The email title must be **R_HW5_GroupName**. NO late homework will be accepted.

P.S.: The total points of HW5 are 110, including 10 bonus points.

Q1. (10%) Import the library *AER* in R, and attach the data set *CPS1988* (like what you did for HW4). Focus on four variables – wage, education, experience, and ethnicity (African-American versus Caucasian). Your job is to finish the following tasks.

(a) Run the linear regression model below (using *lm()*) and save the model as object “CPS_lm”. Note that the ethnicity is a categorical/dummy variable.

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{experience} + \beta_2 \text{education} + \beta_3 \text{experience} * \text{education} + \beta_4 \text{ethnicity} + \varepsilon$$

(b) Run the same model again using *glm()* (family= “gaussian”). Are the results different from the results from part (a)? If not, why is that?

(c) Estimate the same model using *quantile regression* (section 7.5). Similar to what I do in 7.5, run the model from 5% quantile to 95% quantile, and VISUALIZE the impact of each variable on $\log(\text{wage})$ across the whole range of quantiles. Explain what you see.

Q2. (40%) Hand in a short proposal (**1.5 lines spacing & LESS than or equal to two A4 pages**) for your final term project. At a minimum the proposal should 1) motivate the research question, 2) explain the analysis method to use, 3) briefly describe your data (you do not necessarily need data if you decide to do some theoretical modeling work), and 4) clearly state what the objective of the project will be.

Q3. (60%) Did you attend Prof. Chia-Yen Lee’s talk on November 25?