# Economics 217

## Exam #1

**Instructions:** Take home, notes and online resources are fine. You are required to submit a brief write-up showing your results (just like your homework), as well as a source code that is both runnable and readable. Credit will be given based on both materials. Partial credit will be given for code that shows comprehension of the material, even if the end result is incorrect.

### Problem 1

For this question, please use the org data from the class website, we wish to estimate the following model in R代写微信risepaper.

$$rw = s(age) + educ + wbho + gender + u$$

Here, *rw* is the real wage, *educ* is a education factor variable, *wbho* is a factor variable describing the race of the respondent, and *gender* is a factor variable representing gender. You may restrict your model to California in years 2003 and later.

**a.** Please estimate the model described above, using the default smoothing parameters given by the R library that you use. Please display your results using a 2X2 plot, i.e. how each variable affects the outcome. Please label your plots, and interpret your results. (15 points).

**b.** Next, using a K-Nearest Neighbors model, estimate the real wage as a function of age, education, race, and gender. For a white respondent with a college degree, please plot the gender gap - the average difference between male and female wages - as a function of age. (15 Points)

**Extra Credit** Bootstrap your answer in part 'b', and calculate 90% confidence bands for the gender gap. (5 Points)

**Problem 2**

For this question, please use the data set with your name on it from the exam 2 webpage. This dataset was randomly created using the following equation

$$y = \beta_0 + x + \beta_1 \mathbf{1}(x > h_1)(x - h_1) + \beta_2 \mathbf{1}(x > h_2)(x - h_2) + u$$

The variable $x$ is evenly spaced between 0 and 10. The noise parameter $u$ is random from a normal distribution with mean 0 and standard deviation 2. The kink $h_1$ is somewhere in [3, 4] and $h_2$ is somewhere in [7, 8].

Each student has a different vector of kinks, $h_1$ and $h_2$. Your job in these questions is to find the kinks, and provide bootstrap confidence intervals for them.

**a.** Use a leave-one-out (cross validation) procedure to estimate the location of the kinks. Please report the estimates, $\hat{h}_1$ and $\hat{h}_2$, and the estimates for $\beta_0$, $\beta_1$ and $\beta_2$ at these optimally placed kinks. (15 points)

**b.** Since I generated your dataset with noise, I would like to get a sense of the noise in the estimate of the kinks. Please write a bootstrap procedure (resample data, not residuals) to provide a 90% confidence interval for $\hat{h}_1$ and $\hat{h}_2$. (15 points)

**Notes/Hints for this question:**

**Precision:** When finding the optimal kinks, please do so to the nearest tenth (e.g.. 4.1)

**Bootstrap Replications:** Please use at least 20 bootstrap replications for part b. Ideally one would use more, but I want this to be manageable for all computer speeds. On my laptop, using conservative (ie. slow) programming, part *b* takes 15 minutes to run with 20 replications. This could certainly be sped up, and while not required for the question, I do expect that many of you will figure out a way to speed up the procedure. While testing your code, use fewer bootstrap replications to save time waiting.