BUEC-333, Summer 2014
Hand-in assignment #2

Deadline: July 24, 14:30

Rules

1. Use the data that is provided on the course website
2. Use R to answer the questions. Other software is not accepted.
3. Hand in a hardcopy of typed answers, font Times New Roman, font size 12.
4. For **each** question:
   (a) If applicable: copy-paste the code that you used to get that outcome: **without** code, you get **0 points**
   (b) If applicable: copy-paste the R output supporting the answer (this might involve tables and pictures)
   (c) Type **2 or 3 lines** of explanation: **fewer than 2 lines, or more than 3 lines: 0 points**

Unions

Currently, the BC teachers union is on strike. One of the reasons to form a union, and one of the reasons to organize a strike, is to increase your bargaining power in labor negotiations with the employer. It is sometimes believed that unions and strikes improve the wages of the members of that union. In this assignment, you are going to investigate the effects of unions and strikes on wages empirically using linear regression in R. To get started, start R and make sure that the package **foreign** is installed by issuing the command

```
> install.packages("foreign")
```

Now that this package is installed, load the data for our analysis of unions. A description of this data can be found in Table 1 of this paper: [link]

```
> ## Load the "foreign" software
> require(foreign)
> ## Download the data from my website and store it in a data.frame.
```

```
> unionData <- read.dta("http://www.sfu.ca/~cmuris/2014-Summer-333/wagepan.dta")
> ## Delete missing data
> unionData <- na.omit(unionData)
```

To get an idea about the data set, use "str", "summary", "head", etc. Once you are familiar with the data, answer the following questions:

1. Before we start the assignment, enter the following two lines of code and copy the output that R returns into your assignment:

   ```
   > timestamp()
   > getwd()
   ```

2. There is no variable that contains the wage, as only the log of wage is provided. Generate a new variable, "unionData$wage", that contains the hourly wage in US dollars.

3. Make a scatterplot (you can use the package "ggplot", for example) of "wage" and the variable "union", which measures whether somebody is a part of a union or not. What do you conclude from this scatterplot? Does it make a difference when you do it for the log og wage? [Include code and the plot you made.]

4. Run a regression of "wage" on "union". Interpret the regression coefficient estimate for "union". Report the 90% confidence interval for the regression coefficient estimate of union: what do you conclude?

5. Do the same for "lwage".

6. I prefer the model with "lwage" over the model with "wage". However, there is a problem because we did not include any variables other than "union". Explain what the problems are with this, from a theoretical/statistical point of view.

7. Run a regression of "lwage" on union, hours, year, occ1, occ2, occ3, occ4, occ5, occ6, occ7, occ8, occ9. R refuses to give an estimate for the regression coefficient on occ9. Why?

8. Interpret the estimate of the regression coefficient of "year".

9. I wonder whether I should include a person's work experience, as measured by the variable "exper". Try it, and argue why it should or should not be included. You can be informal.

10. The relationship between wages and experience seems suspicious. There may still be an omitted variable problem. Include "educ". What do you conclude about education and experience, and their relationship with wages?

11. Are unions good for employees' wages?