

Assignment 4

Use the data “injury.dta” for this assignment. You can download the data from the following link:

https://www.dropbox.com/sh/4hvu3aevi3zifia/AABrIgRuYLO5RfMJ4VB_z6O

The goal of this exercise is to evaluate the impact of compensation for injuries at work on injury duration. Look at the variable description by using the following code to get their exact definition.

```
library(foreign)
myData=read.dta("e:/INJURY.dta")
var.labels = attr(myData,"var.labels")
data.frame(var.name=names(myData),var.labels)
```

Q1. What is the expected effect of compensation on the injury duration? Can we do a simple regression of the compensation on the duration? Why or why not?

Q2. Explain that there could be different responses for low-earners or high-earners.

Q3. Regress the log of duration (durat) on a dummy indicating whether the observation is before or after the change in benefits (afchnge), a dummy indicating whether the individual is a high earner or not (highearn), and the interaction of both (afchnge*highearn). Interpret the results. **DDD**

Q4. How could omitted variables affect the estimation? Include the control variables that seem adequate in your regression model. Try to justify why we need those variables. Compare the results of the model with control variables and the model without control variables.

Q5. Explain the potential problem of the above model.

Guideline for Assignment 4:

Submit your answer and R-codes used for the analysis to the course Blackboard. Please include your student number and name in the header of the document. Your document should be in the same format as the final report (See the syllabus) and should not exceed three A4 pages. Append R-codes used for the analysis. There is no page limit for the appendix. The due is **May 6, 23:59 PM**.