From:	John McConnell
To:	Aaron Mendelson; Benjamin Chan; Hyunjee Kim; Jonah Todd-Geddes; Lorie Jacob; Margaret Sarna-Wojcicki; Peter Graven; Ruth Rowland; Stephanie Renfro
Subject:	Developing programming examples for R, Stata, SAS
Date:	Tuesday, December 03, 2013 9:18:00 AM

Group

Following up on our discussion last week, I propose the following exercise. I have tried to keep this very simple. We can get more involved if we think that would help. This is a baby step.

Let's have these people take the lead:

Stata – Peter
R – Stephanie
SAS – Ben

Let's work off this directory
X:\OHSU Shared\Restricted\OCHSER\PROJECTS\Code_Sharing
Using the file "Inpatient_Claims_Sample_1." (I have SAS, Stata, and R versions there).

Please create a word document and a statistical code file (e.g., a Stata "do" file or analog in SAS or R). The word document should describe, briefly

a. How you start the program
b. How you locate the file Inpatient_Claims_Sample_1 within your program (or perhaps reference this in the code file).
c. How you run your code
d. How you view your results

It is ok if the word file is only 4 sentences! I prefer to err on the side of keeping it simple at this point. We can always go back and add more detail later. This is not a manual.

The code file should be commented and describe how to do the following

a. Load the dataset
b. Describe the variables that exist in the data set (i.e., print out the names of the variables)
c. Take the mean of CLM_PMT_AMT
d. Run an OLS regression of CM_PMT_AMT on CLM_UTLZTN_DAY_CNT and CLM_PASS_THRU_PER_DIEM_AMT (coefficient on CLM_UTLZTN_DAY_CNT should be about 489 and R-squared should be about 0.11)
e. Save the output (the coefficient estimates, standard errors, p-values) in a text file or other file that can be opened up easily in Word or Notepad or a browser.

When you're ready, send your Word doc and Code file to the group

WHEN?

By January 3, 2014

<u>Ben, Peter, Stephanie: let me know by the end of the week if you'd like to change this plan (I.e., if it seems confusing, useless, or you're too busy).</u>

Also, Ben made the following observations, which I agree with – but I thought they might add to the complexity and so I thought we could revisit these at some point in the future, and for now just keep it very very simple.

COMMENTS FROM BEN

I think it might actually be instructive to see how to get the data from CSV to a native format for analysis (how easy/hard it is, how much can be automated, how much needs to be programmed). That might be a big enough task just to make it the first "project".

We'll want to put data on the X-drive. Probably a subfolder in the PROJECTS folder called "CodeSharing" or something like that and further organized with the following folders

1. *Data*
2. *Task01ImportingData*
3. *Task02ComputingUnivariateStats*
4. *Task03ModelingOLS*
5. *Etc*